

# Anisotropic Nonlocal Means Denoising

Arian Maleki\*, Manjari Narayan, Richard G. Baraniuk<sup>1,\*\*</sup>

*Dept. of Computer and Electrical Engineering, Rice University, MS-380  
6100 Main Street, Houston, TX 77005, USA*

---

## Abstract

It has recently been proved that the popular nonlocal means (NLM) denoising algorithm does not optimally denoise images with sharp edges. Its weakness lies in the isotropic nature of the neighborhoods it uses to set its smoothing weights. In response, in this paper we introduce several theoretical and practical *anisotropic nonlocal means* (ANLM) algorithms and prove that they are near minimax optimal for edge-dominated images from the Horizon class. On real-world test images, an ANLM algorithm that adapts to the underlying image gradients outperforms NLM by a significant margin.

*Keywords:* Denoising, nonlocal means, minimax risk, anisotropy

---

## 1. Introduction

Image denoising is a fundamental primitive in image processing and computer vision. Denoising algorithms have evolved from the classical linear and median filters to more modern schemes like total variation denoising [51], wavelet thresholding [28], and bilateral filters [21, 53, 57, 66].

---

\*Corresponding author

\*\*Principal corresponding author

*Email addresses:* [arian.maleki@rice.edu](mailto:arian.maleki@rice.edu) (Arian Maleki), [manjari@rice.edu](mailto:manjari@rice.edu) (Manjari Narayan), [richb@rice.edu](mailto:richb@rice.edu) (Richard G. Baraniuk)

*URL:* <http://www.ece.rice.edu/~mam15/> (Arian Maleki),  
<http://www.ece.rice.edu/~mn4/> (Manjari Narayan),  
<http://web.ece.rice.edu/richb/> (Richard G. Baraniuk)

<sup>1</sup>Phone: +1 713.348.5132; Fax: +1 713.348.5685

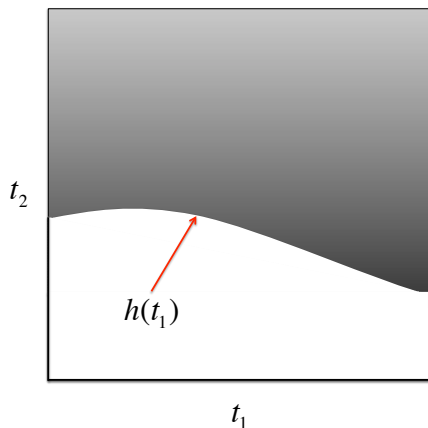


Figure 1: An example of a Horizon class image that features a smooth edge contour that separates the white region from the black region.

A particularly successful denoising scheme is the *nonlocal means* (NLM) algorithm [8], which estimates each pixel value as a weighted average of other, similar noisy pixels. However, instead of using spatial adjacency or noisy pixel value as the similarity measure to adjust the estimate weights, NLM uses a more reliable notion of similarity based on the resemblance of the pixels' neighborhoods in high-dimensional space. This unique feature benefits NLM in two ways. First, it provides more accurate weight estimates. Second, it enables NLM to exploit the contribution of all pixels in the image. In concert, these features enable NLM to provide remarkable performance for a large class of image denoising problems.

Nevertheless, in a recent paper, we have proved that NLM does not attain optimal performance on images with sharp edges from the so-called *Horizon class* (see Figure 1) [41].<sup>2</sup> Indeed, NLM's theoretical performance is more or less equivalent to wavelet thresholding, which was shown to be suboptimal in [17]. Empirical results have also confirmed the suboptimality of NLM in estimating sharp edges [14, 15, 18, 52]. The core problem is that NLM (and wavelet thresholding) cannot exploit the smoothness of the edge contour that separates the white and black regions.

In this paper, we introduce and study a new denoising framework and

---

<sup>2</sup>Recently [5] has extended our results to more complicated image/edge models.

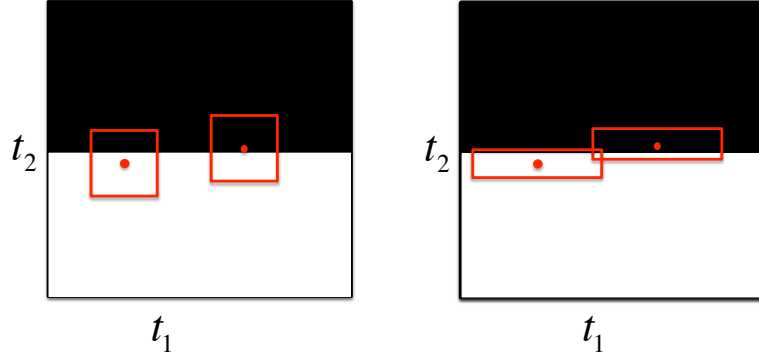


Figure 2: Comparison of (left) the isotropic neighborhoods employed by Non Local Means (NLM) versus (right) anisotropic neighborhoods employed by Anisotropic NLM (ANLM).

prove that it is near-optimal for Horizon class images with sharp edges. *Anisotropic nonlocal means* (ANLM) outperforms NLM, wavelet thresholding, and more classical techniques by using anisotropic neighborhoods that are elongated along and matched to the local edge orientation. Figure 2 compares the neighborhoods used in ANLM with those used in NLM. Anisotropic neighborhoods enable ANLM to distinguish between similar and dissimilar pixels more accurately.

We develop three different ANLM algorithms of increasing levels of practicality. *Oracle Anisotropic Nonlocal Means* (OANLM) assumes perfect knowledge of the local orientation of the edge contour and is used primarily for our theoretical optimality analysis. *Discrete-angle Anisotropic Nonlocal Means* (DANLM) optimizes the choice of the anisotropic neighborhood around each pixel in order to achieve near-optimal performance without any oracle information. Since it is more computationally demanding than NLM, we introduce an algorithmic simplification. *Gradient based Anisotropic Nonlocal Means* (GANLM) uses image gradient information to estimate the edge orientation; using simulations, we demonstrate that GANLM significantly outperforms NLM in practice on both Horizon class and real-world images.

Anisotropy is a fundamental feature of real-world images. Hence, many denoising algorithms have been proposed to exploit this feature [9, 11, 14–16, 35, 44, 46, 47, 52, 55, 56, 61, 63, 64]. We will review all these algorithms

and their similarities and differences with ANLM in Section 5.

The paper is organized as follows. Section 2 explains the minimax framework we use to analyze the denoising algorithms and reviews the necessary background. Section 3 introduces the OANLM and DANLM algorithms and presents the main theorems. Section 4 addresses some of the practical ANLM issues by introducing GANLM and summarizes the results of a range of simulations using synthetic and real-world imagery. Section 5 reviews the related work in the literature. Section 6 discusses our current and potential future results. Section 7 contains the proofs of the main theorems.

## 2. Minimax analysis framework

In this section we introduce the minimax framework [28, 37] and the Horizon class image model considered in this paper. Note that, in order to streamline the proofs, we take a continuous-variable analysis approach in this paper, in contrast to our approach in [41]. The moral of the story is compatible with [41], however.

### 2.1. Risk

We are interested in estimating an image described by the function  $f : [0, 1]^2 \rightarrow [0, 1]$  ( $f \in L^2([0, 1]^2)$ ) from its noisy observation

$$dY(t_1, t_2) = f(t_1, t_2)dt_1dt_2 + \sigma dW(t_1, t_2). \quad (1)$$

Without loss of generality, we consider only square images. Here  $W(t_1, t_2)$  is the Wiener sheet,<sup>3</sup> and  $\sigma$  is a constant that scales the noise. For a given function  $f$  and a given estimator  $\hat{f}$ , define the *risk* as

$$R(f, \hat{f}) = \mathbb{E}(\|f - \hat{f}\|_2^2),$$

where the expected value is over  $W$ . The risk can be decomposed into bias squared and variance terms

$$R(f, \hat{f}) = \|f - \mathbb{E}\hat{f}\|_2^2 + \mathbb{E}\|\hat{f} - \mathbb{E}\hat{f}\|_2^2.$$

---

<sup>3</sup>The Wiener sheet is the primitive of white noise.

Let  $f$  belong to a class of functions  $\mathcal{F}$ . The risk of the estimator  $\hat{f}$  on  $\mathcal{F}$  is defined as the risk of the least-favorable function, i.e.,

$$R(\mathcal{F}, \hat{f}) = \sup_{f \in \mathcal{F}} R(f, \hat{f}).$$

The *minimax* risk over the class of functions  $\mathcal{F}$  is then defined as

$$R^*(\mathcal{F}) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} R(f, \hat{f}).$$

$R^*(\mathcal{F})$  is a lower bound for the performance of any estimator on  $\mathcal{F}$ .

In this paper, we are interested in the asymptotic setting as  $\sigma \rightarrow 0$ . For all the estimators we consider,  $R(\mathcal{F}, \hat{f}) \rightarrow 0$  as  $\sigma \rightarrow 0$ . Therefore, following [10, 48] we will consider the *decay rate of the minimax risk* as our measure of performance. For this purpose, we will use the following asymptotic notation.

**Definition 1.**  $f(\sigma) = O(g(\sigma))$  as  $\sigma \rightarrow 0$ , if and only if there exist  $\sigma_0$  and  $c$  such that for any  $\sigma < \sigma_0$ ,  $|f(\sigma)| \leq c|g(\sigma)|$ . Likewise,  $f(\sigma) = \Omega(g(\sigma))$  as  $\sigma \rightarrow 0$ , if and only if there exist  $\sigma_0$  and  $c$  such that for any  $\sigma < \sigma_0$ ,  $|f(\sigma)| \geq c|g(\sigma)|$ . Finally,  $f(\sigma) = \Theta(g(\sigma))$ , if and only if  $f(\sigma) = O(g(\sigma))$  and  $f(\sigma) = \Omega(g(\sigma))$ . We may interchangeably use  $f(\sigma) \asymp g(\sigma)$  for  $f(\sigma) = \Theta(g(\sigma))$ .

**Definition 2.**  $f(\sigma) = o(g(\sigma))$  if and only if  $\lim_{\sigma \rightarrow 0} \frac{f(\sigma)}{g(\sigma)} = 0$ .

## 2.2. Horizon edge model

In our analysis, we consider the *Horizon* model that contains piecewise constant images with sharp step edges lying along a smooth contour [17, 34, 48] (our analysis extends easily to piecewise smooth edges). Let  $\text{Hölder}^\alpha(C)$  be the class of Hölder functions on  $\mathbb{R}$ , defined in the following way:  $h \in \text{Hölder}^\alpha(C)$  if and only if

$$|h^{(k)}(t_1) - h^{(k)}(t'_1)| \leq C|t_1 - t'_1|^{\alpha-k},$$

where  $k = \lfloor \alpha \rfloor$ . Consider a transformation that maps each one-dimensional edge contour function  $h$  to a two-dimensional image  $f_h : [0, 1]^2 \rightarrow [0, 1]$  via

$$f_h(t_1, t_2) = \mathbf{1}_{\{t_2 < h(t_1)\}}.$$

The Horizon class of images is then defined as

$$H^\alpha(C) = \{f_h(t_1, t_2) : h \in \text{Hölder}^\alpha(C) \cap \text{Hölder}^1(1)\}, \quad (2)$$

where  $\alpha$  is the smoothness of the edge contour. Figure 1 illustrates a sample function of this class. The following theorem characterizes the minimax rate of  $H^\alpha(C)$  [17, 34, 48].<sup>4</sup>

**Theorem 1.** [17, 34, 48] *For  $\alpha \geq 1$ , the minimax risk of the class  $H^\alpha(C)$  is*

$$R^*(H^\alpha(C)) = \Theta\left(\sigma^{\frac{2\alpha}{\alpha+1}}\right).$$

Achieving the minimax rate is a laudable goal that any well-respecting denoising algorithm should aspire to. In this paper, we will focus primarily on  $\alpha = 2$  edge contours, for which the optimal minimax decay rate is  $\sigma^{4/3}$ . However, it is straightforward to draw similar conclusions for the other values of  $\alpha$ .

### 2.3. Image denoising algorithms

Minimax risk analysis of the classical denoising algorithms has revealed their suboptimal performance on images with sharp edges. Table 1 summarizes several of these algorithms and their minimax decay rates (up to a log factor).<sup>5</sup> The suboptimality of wavelet thresholding has led to the development of ridgelets [11], curvelets [9], wedgelets [17], platelets [63], shearlets [35], contourlets [16], bandelets [46], directionlets [61] and other types of directional transforms. See [27, 42] and the references therein for more information. In the framework of this paper, wedgelet denoising [17] provably achieves the optimal minimax rate. However, it performs poorly on textures, which has limited its application in image processing.

---

<sup>4</sup>The models considered in [17, 34, 48] are slightly different from the continuous framework of this paper. Therefore, for the sake of completeness we prove Theorem 1 in Appendix C.

<sup>5</sup>The analysis framework used in [4] and [41] is discrete rather than continuous. Therefore, for the sake of completeness we establish the results for the mean filter and NLMM in Appendix A and Appendix B, respectively.

Table 1: Minimax risk decay rates of several classical image denoising algorithms; recall that the according to Theorem 1, for  $\alpha = 2$  the optimal minimax rate is  $\sigma^{4/3}$ .

Algorithm	Minimax Rate
Mean filter [4]	$\sigma^{2/3}$
Wavelet thresholding [17, 48]	$\sigma^1$
Nonlocal means [41]	$\sigma^1$

#### 2.4. Nonlocal means denoising

In 2005, Buades, Coll, and Morel significantly improved the performance of the bilateral filter [57] by incorporating a new notion of pixel similarity. As in the mean filter, NLM estimates each pixel value using a weighted average of other pixel values in the image. However, NLM sets the weights according to the similarity between the pixel neighborhoods rather than the pixel values. Furthermore, in contrast to the bilateral filter, in which only the vicinity of each pixel contributes to the estimate, in NLM all pixels may contribute.

Specifically, the NLM algorithm is defined as follows. Let  $S \triangleq [0, 1]^2$  represent the domain of the image. Define the  $\delta$ -neighborhood of a  $(t_1, t_2)$  as

$$I_\delta(t_1, t_2) \triangleq \left[ t_1 - \frac{\delta}{2}, t_1 + \frac{\delta}{2} \right] \times \left[ t_2 - \frac{\delta}{2}, t_2 + \frac{\delta}{2} \right]. \quad (3)$$

Following the definition of NLM in the discrete setting, we pixelate this neighborhood by partitioning  $I_\delta(t_1, t_2)$  into  $n^2$  subregions  $I_\delta^{j_1, j_2} \triangleq [t_1 - \frac{j_1\delta}{2n}, t_1 + \frac{j_1\delta}{2n}] \times [t_2 - \frac{j_2\delta}{2n}, t_2 + \frac{j_2\delta}{2n}]$ . The pixelated neighborhood of  $(t_1, t_2)$  is defined as  $\mathbf{y}_{t_1, t_2}^\delta \in \mathbb{R}^{n \times n}$  and satisfies

$$\mathbf{y}_{t_1, t_2}^\delta(j_1, j_2) \triangleq \frac{n^2}{\delta^2} \int_{(s_1, s_2) \in I_\delta^{j_1, j_2}} dY(s_1, s_2).$$

We further define the pixelated process as

$$X(t_1, t_2) \triangleq \frac{n^2}{\delta^2} \int_{(s_1, s_2) \in I_{\delta/n}(t_1, t_2)} dY(s_1, s_2).$$

Define the  $\delta$ -neighborhood distance between two points in the image as

$$d_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) \triangleq \frac{1}{n^2 - 1} (\|\mathbf{y}_{t_1, t_2}^\delta - \mathbf{y}_{s_1, s_2}^\delta\|_2^2 - |\mathbf{y}_{t_1, t_2}^\delta(0, 0) - \mathbf{y}_{s_1, s_2}^\delta(0, 0)|^2), \quad (4)$$

where  $\|\mathbf{y}_{t_1, t_2}^\delta\|_2^2 \triangleq \sum_{i,j} (\mathbf{y}_{t_1, t_2}^\delta(i, j))^2$ . Note that in contrast to the definition in [8], we have removed the center element  $|\mathbf{y}_{t_1, t_2}^\delta(0, 0) - \mathbf{y}_{s_1, s_2}^\delta(0, 0)|^2$  from the summation. As we will see in Section 3,  $n^2 \rightarrow \infty$  as  $\sigma \rightarrow 0$ , and hence the effect of removing the center pixel is negligible on the asymptotic performance. But, as we will see below, removing the center element simplifies the calculations considerably. NLM uses the neighborhood distances to estimate

$$\hat{f}^N(t_1, t_2) \triangleq \frac{\int_{(s_1, s_2) \in S} w_{t_1, t_2}^N(s_1, s_2) X(s_1, s_2) ds_1 ds_2}{\int_{(s_1, s_2) \in S} w_{t_1, t_2}^N(s_1, s_2) ds_1 ds_2},$$

where  $w_{t_1, t_2}^N(s_1, s_2)$  is set according to the  $\delta$ -neighborhood distance between  $dY(t_1, t_2)$  and  $dY(s_1, s_2)$ .<sup>6</sup> It is straightforward to verify that  $\mathbb{E} d_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) = d_\delta^2(f(t_1, t_2), f(s_1, s_2)) + \frac{2n^2\sigma^2}{\delta^2}$ , which suggests the following strategy for setting the weights:

$$w_{t_1, t_2}^N(s_1, s_2) \triangleq \begin{cases} 1 & \text{if } d_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) \leq \frac{2n^2\sigma^2}{\delta^2} + \tau_\sigma, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where  $\tau_\sigma$  is the *threshold parameter*. Soft/tapered weights have been explored and are often used in practice [8]. However, the above untapered weights capture the essence of the algorithm while simplifying the analysis.

The distinguishing feature of NLM — the weighted averaging of pixels based on the neighborhoods — produces a decay rate that is superior to that of linear filters as shown in [41]. However, compared to the optimal rate of  $\sigma^{4/3}$ , NLM remains suboptimal. We introduce the anisotropic NLM algorithm in the next section to address this gap in performance.

### 3. Anisotropic nonlocal means denoising

In this section, we introduce the Anisotropic Nonlocal Means (ANLM) concept that exploits the smoothness of edge contours via anisotropic neigh-

---

<sup>6</sup>We assume that both  $w_{t_1, t_2}^N(s_1, s_2) X(s_1, s_2)$  and  $w_{t_1, t_2}^N(s_1, s_2)$  are Lebesgue integrable with high probability.



borhoods. We then present OANLM and DANLM algorithms and explain their performance guarantees.

### 3.1. Directional neighborhoods

We now formally introduce the notion of directional neighborhoods. Then we extend NLM to exploit such neighborhoods. This will lay the foundation for the OANLM and DANLM algorithms.

Let  $R_{\nu,\mu}^\theta(\cdot)$  represent the rotation operator on a neighborhood; when applied to a generic point  $(u, v) \in S$  in the neighborhood around  $(\nu, \mu)$ ,  $R_{\nu,\mu}^\theta$  rotates  $(u, v)$  by  $\theta^\circ$  counter-clockwise around the point  $(\nu, \mu)$ . For a set  $Q \subset S$  we define  $Q_\theta \triangleq R_{\nu,\mu}^\theta(Q)$  as

$$(s, t) \in Q_\theta \Leftrightarrow \exists (u, v) \in Q \text{ such that } (s, t) = R_{\nu,\mu}^\theta((u, v)).$$

The  $(\theta, \delta_s, \delta_\ell)$ -anisotropic neighborhood of the point  $(t_1, t_2)$  is defined as

$$I_{\theta, \delta_s, \delta_\ell}(t_1, t_2) \triangleq R_{t_1, t_2}^\theta \left( \left[ t_1 - \frac{\delta_\ell}{2}, t_1 + \frac{\delta_\ell}{2} \right] \times \left[ t_2 - \frac{\delta_s}{2}, t_2 + \frac{\delta_s}{2} \right] \right) \cap S,$$

where  $\theta$ ,  $\delta_\ell$ , and  $\delta_s$  ( $\delta_s \leq \delta_\ell$ ) represent the orientation angle, length, and width of the neighborhood, respectively. Figure 3 displays such neighborhoods for two different pixels. We partition  $I_{\theta, \delta_s, \delta_\ell}(t_1, t_2)$  into  $n_s \times n_\ell$  subregions  $I_{\theta, \delta_s, \delta_\ell}^{j_1, j_2}(t_1, t_2) \triangleq R_{t_1, t_2}^\theta \left( \left[ t_1 - \frac{j_1 \delta_\ell}{2n_\ell}, t_1 + \frac{j_1 \delta_\ell}{2n_\ell} \right] \times \left[ t_2 - \frac{j_2 \delta_s}{2n_s}, t_2 + \frac{j_2 \delta_s}{2n_s} \right] \right)$ .

The pixelated neighborhood of  $(t_1, t_2)$  is defined as the  $\mathbf{y} \in \mathbb{R}^{n_s \times n_\ell}$  satisfying

$$\mathbf{y}_{t_1, t_2}^{\theta, \delta_\ell, \delta_s}(j_1, j_2) \triangleq \frac{n_s n_\ell}{\delta_s \delta_\ell} \int \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_\ell}^{j_1, j_2}(t_1, t_2)} dY(s_1, s_2).$$

Define the neighborhood process as

$$X(t_1, t_2) \triangleq \frac{n_s n_\ell}{\delta_s \delta_\ell} \int \int_{(s_1, s_2) \in I_{\theta, \frac{\delta_s}{n_s}, \frac{\delta_\ell}{n_\ell}}(t_1, t_2)} dY(s_1, s_2).$$

The anisotropic  $(\delta_\ell, \delta_s, \theta)$ -neighborhood distance between two given points in the image is then defined as

$$\begin{aligned} & d_{\theta, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(s_1, s_2)) \\ & \triangleq \frac{1}{n_s n_\ell - 1} \left( \|\mathbf{y}_{t_1, t_2}^{\theta, \delta_\ell, \delta_s} - \mathbf{y}_{s_1, s_2}^{\theta, \delta_\ell, \delta_s}\|_2^2 - |\mathbf{y}_{t_1, t_2}^{\theta, \delta_\ell, \delta_s}(0, 0) - \mathbf{y}_{s_1, s_2}^{\theta, \delta_\ell, \delta_s}(0, 0)|^2 \right). \quad (6) \end{aligned}$$

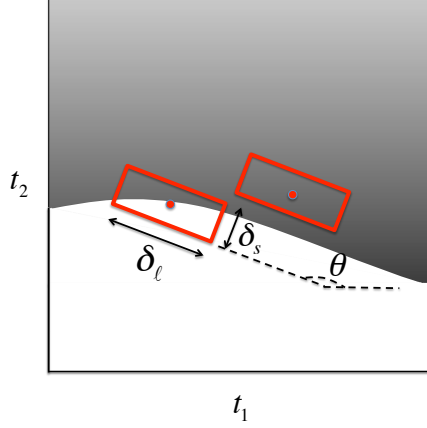


Figure 3: The anisotropic neighborhood  $I_{\theta, \delta_s, \delta_\ell}(t_1, t_2)$  in the discrete setting for two different pixels of a Horizon class image.

The ANLM estimate at point  $(t_1, t_2)$  is given by

$$\hat{f}^{\theta, \delta_s, \delta_\ell}(t_1, t_2) \triangleq \frac{\int_{(s_1, s_2) \in S} w_{t_1, t_2}^{\theta, \delta_s, \delta_\ell}(s_1, s_2) X(s_1, s_2) ds_1 ds_2}{\int \int_{(s_1, s_2) \in S} w_{t_1, t_2}^{\theta, \delta_s, \delta_\ell}(s_1, s_2) ds_1 ds_2}.$$

where the weights are obtained from

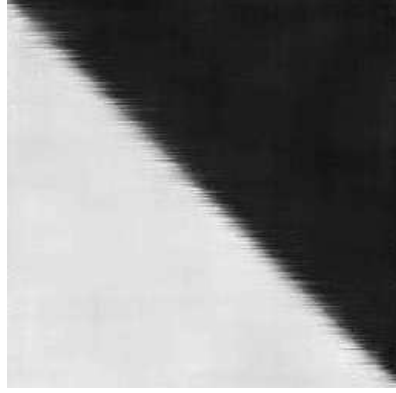
$$w_{t_1, t_2}^{\theta, \delta_s, \delta_\ell}(s_1, s_2) = \begin{cases} 1 & \text{if } d_{\theta, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(s_1, s_2)) \leq \frac{2n_s n_\ell \sigma^2}{\delta_s \delta_\ell} + \tau_\sigma, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Here  $\tau_\sigma$  is the *threshold parameter* and  $\frac{2n_s n_\ell \sigma^2}{\delta_s \delta_\ell} + \tau_\sigma$  is the *threshold value*.

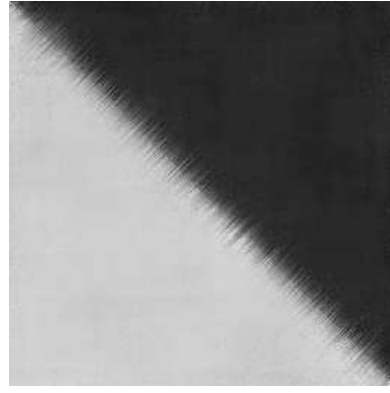
ANLM extends NLM in two significant ways. First, in ANLM  $\theta, \delta_s, \delta_\ell$  are free parameters, while in NLM  $\theta = 0$  and  $\delta_s = \delta_\ell$ . As Figures 4 and 5 demonstrate, these free parameters significantly affect the performance of ANLM. Figure 4 illustrates the visual denoising results as we vary the angle of the anisotropic neighborhood. Figure 5 does the same for the empirical peak signal-to-noise ratio (PSNR).<sup>7</sup>

---

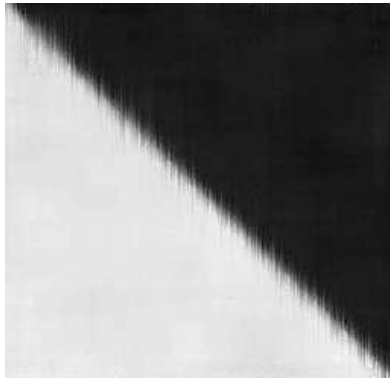
<sup>7</sup>In practice images are measured on a discrete lattice, i.e., the pixelated values are known. Let the pixelated values of the original noise-free image and the estimator be



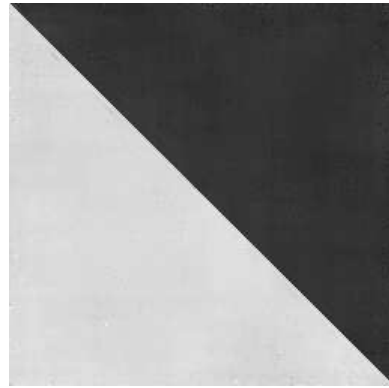
(a)  $0^\circ$ , PSNR = 20.6dB



(b)  $45^\circ$ , PSNR = 19.2dB



(c)  $90^\circ$ , PSNR = 20.6dB



(d)  $135^\circ$ , PSNR = 28.9dB

Figure 4: Effect of the anisotropic neighborhood orientation angle  $\theta$  on the denoising performance of ANLM. The images are  $256 \times 256$  pixels with a true edge orientation of  $135^\circ$ . The noise standard deviation is  $\xi = 0.5$ , which is equivalent to PSNR = 6.1dB (see Section 4.1). The neighborhood size and threshold parameter are  $\delta_s = 1/256$ ,  $\delta_\ell = 20/256$  and  $\tau = 2.7\xi^2$ , respectively. The angles of the neighborhoods ( $\theta$ ) and the resulting PSNR of the estimates are noted below the images. There is a substantial subjective and objective improvement in the performance of ANLM when the neighborhood is aligned with the edge.

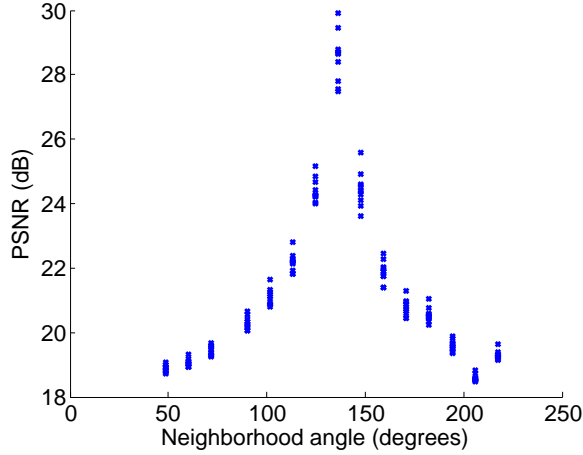


Figure 5: Effect of the anisotropic neighborhood orientation angle  $\theta$  on the denoising performance of ANLM. The experimental parameters are the same as in Figure 4. For each value of  $\theta$  we plot the result of 10 Monte Carlo simulations. PSNR of the original noisy image is 6.1dB

### 3.2. Oracle ANLM

It is clear from Figures 4 and 5 that we should align the ANLM neighborhood locally with the edge to maximize denoising performance. In the Oracle ANLM (OANLM) algorithm, we assume that we have access to an oracle that knows the image's edge orientation at each pixel. Consider  $f_h(t_1, t_2) \in H^\alpha(C)$ . Let  $h'(t_1)$  denote the derivative of the edge contour  $h(t_1)$ , and let  $\Gamma_\gamma \triangleq \{(t_1, t_2) : t_1 = \gamma\}$  segment the region  $S$ . Define  $\theta_\gamma$  as the angle between the tangent to the edge contour  $h$  and the horizontal line at  $t_1 = \gamma$ . OANLM is defined as ANLM with the following parameter settings:

- *Quadratic scaling:* Instead of using isotropic (square) neighborhoods, we use anisotropic (rectangular) neighborhoods of size  $\delta_s \times \delta_\ell$ . The scaling of  $\delta_s$  and  $\delta_\ell$  depends on the smoothness of the edge. Since

---

$f_{ij}$  and  $\hat{f}_{i,j}$ , respectively. The empirical mean-squared-error (MSE) of an estimator  $\hat{f}$  is defined as  $\text{MSE} = \frac{1}{n^2} \sum_{i,j} (\hat{f}_{i,j} - f_{i,j})^2$ . If  $f_{i,j} \in [0, A]$ , then the empirical peak signal-to-noise ratio (PSNR) is defined as  $10 \log_{10} \frac{A^2}{\text{MSE}}$ .

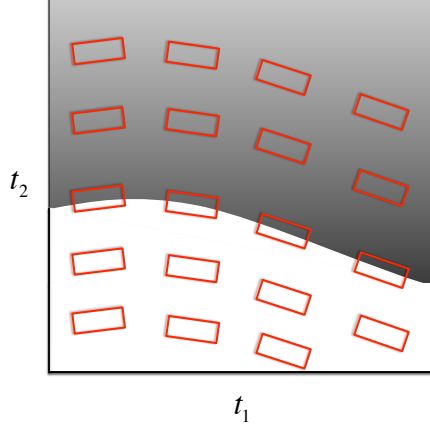


Figure 6: The OANLM neighborhoods satisfy the quadratic scaling  $\delta_s = \delta_\ell^2$  and are aligned with the edge. In regions far from the edge, isotropic neighborhoods perform just as well.

we are primarily interested in  $H^2(C)$ , we will emphasize the quadratic scaling<sup>8</sup>  $\delta_s = 4\sigma^{4/3}|\log \sigma|^{4/3}$  and  $\delta_\ell = 2\sigma^{2/3}|\log \sigma|^{2/3}$ ; that is  $\delta_s = \delta_\ell^2$ .

- *Aligned neighborhoods:* We align the neighborhood at point  $(t_1, t_2)$  to the orientation  $\theta_{t_1}$  of the edge contour at  $(t_1, h(t_1))$ . Thus all pixels in a column have the same orientation. Figure 6 illustrates such a neighborhood selection.<sup>9</sup>
- *Logarithmic pixelation:* We assume that  $n_s \times n_\ell = \lceil \log^2 \sigma \rceil$ . In other words, as  $\sigma$  increases the resolution of the pixelated image increases as well.

**Theorem 2.** If  $\hat{f}^O$  is the OANLM estimator with the threshold  $\tau_\sigma = \frac{2}{\sqrt{|\log \sigma|}}$ , then

$$R(H^\alpha(C), \hat{f}^O) = O(\sigma^{4/3}|\log \sigma|^{4/3}).$$

<sup>8</sup>If the smoothness of an edge is  $\alpha > 1$ , then the optimal neighborhood size is given by  $\delta_s = \Theta(\sigma^{2\alpha/(\alpha+1)}|\log \sigma|^{2\alpha/(\alpha+1)})$  and  $\delta_\ell = \Theta(\sigma^{2/(\alpha+1)}|\log \sigma|^{2/(\alpha+1)})$ . Note that for the special case  $\alpha = 2$  they satisfy the quadratic scaling.

<sup>9</sup>In smooth regions, where the pixel neighborhoods do not intersect with the edge, neither the shape nor the orientation of the neighborhood affect the denoising performance.

The proof of Theorem 2 can be simply modified to provide a bound for the risk of NLM as well. Let  $\delta = \sqrt{\delta_s \delta_\ell}$  and  $n = \sqrt{n_s n_\ell} = |\log(\sigma)|$ , where  $\delta_s, \delta_\ell, n_s, n_\ell$  are as given in OANLM algorithm.

**Corollary 1.** *If  $\hat{f}^N$  is the NLM estimator with threshold  $\tau_\sigma = 2/\sqrt{|\log(\sigma)|}$ , then*

$$R(H^\alpha(C), \hat{f}^N) = O(\sigma |\log \sigma|).$$

See Section 7.3 for the proof of this corollary. Also, under certain mild assumptions, the risk of NLM can be lower bounded by  $\Omega(\sigma)$ . The proof is similar to the proof of Theorem 5 in [41]. However, for the sake of completeness and since our framework differs from [41], we overview the main steps of the proof in Appendix B.

Theorem 2 is based on the strong oracle assumption that the edge direction is known exactly. Needless to say, such information is rarely available in practical applications. Consider a weaker notion of OANLM that has access to an estimate  $\hat{\theta}_\gamma$  of  $\theta_\gamma$  that satisfies

$$|\hat{\theta}_\gamma - \theta_\gamma| \leq \Theta(\sigma^\beta). \quad (8)$$

OANLM with exact edge orientation information corresponds to  $\beta = \infty$  in this model. When  $\beta < \infty$ , the choices  $\delta_\ell = 2\sigma^{2/3} |\log \sigma|^{2/3}$  and  $\delta_s = 4\sigma^{4/3} |\log \sigma|^{4/3}$  are not necessarily optimal. The following theorem characterizes the performance of OANLM using  $\hat{\theta}$ .

**Theorem 3.** *Let the OANLM estimator use the inaccurate edge orientation  $\hat{\theta}_\gamma$  satisfying (8). Set the neighborhood sizes to  $\delta_\ell = \min(\sigma^{2/3} |\log^{2/3} \sigma|, \sigma^{1-\beta/2} |\log \sigma|)$  and  $\delta_s = \sigma^2 \log^2 \sigma / \delta_\ell$ . Then the risk of the estimator with  $\tau_\sigma = 2/\sqrt{|\log(\sigma)|}$  satisfies*

$$R(H^\alpha(C), \hat{f}^O) = O(\delta_s \text{Poly}(|\log \sigma|)),$$

where  $\text{Poly}(|\log \sigma|)$  is a polynomial of degree at most 2 in terms of  $|\log \sigma|$ .

If the edge estimate is exact, then the result simplifies to the result of Theorem 2. However, this theorem confirms that OANLM algorithm achieves the optimal rate even if there is an error in  $\hat{\theta}$  of order  $O(\sigma^\beta)$  with  $\beta \geq 2/3$ .

**Corollary 2.** *Let the OANLM estimator use the inaccurate edge orientation  $\hat{\theta}_\gamma$  satisfying (8) with  $\beta \geq 2/3$ . Set the neighborhood sizes to  $\delta_\ell = 2\sigma^{2/3}|\log^{2/3} \sigma|$  and  $\delta_s = 4\sigma^{4/3}|\log \sigma|$ . Then, the risk of the estimator with  $\tau_\sigma = 2/\sqrt{|\log(\sigma)|}$  satisfies*

$$R(H^\alpha(C), \hat{f}^O) = O(\sigma^{4/3}\text{Poly}(|\log \sigma|)),$$

where  $\text{Poly}(|\log \sigma|)$  is a polynomial of degree at most 4 in terms of  $|\log(\sigma)|$ .

*Proof.* If we plug in  $\beta \geq 2/3$  in Theorem 3 we obtain

$$\delta_\ell = \min(2\sigma^{2/3}|\log^{2/3} \sigma|, 2\sigma^{1-\beta/2}|\log \sigma|) = 2\sigma^{2/3}|\log \sigma|^{2/3}.$$

Therefore,  $\delta_s = 4\sigma^{4/3}|\log \sigma|^{4/3}$ . This establishes the result.  $\square$

We can also simplify the result of Theorem 3 for the case of  $\beta < 2/3$ .

**Corollary 3.** *Let the OANLM estimator use the inaccurate edge orientation  $\hat{\theta}_\gamma$  satisfying (8) with  $\beta \leq 2/3$ . Set the neighborhood sizes to  $\delta_\ell = 2\sigma^{1-\beta/2}|\log \sigma|$  and  $\delta_s = 4\sigma^{1+\beta/2}|\log \sigma|$ . Then, the risk of the estimator with  $\tau_\sigma = 2/\sqrt{|\log(\sigma)|}$  satisfies*

$$R(H^\alpha(C), \hat{f}^O) = O(\sigma^{1+\beta/2}\text{Poly}(|\log \sigma|)).$$

where  $\text{Poly}(|\log \sigma|)$  is a polynomial of degree at most 3 in terms of  $|\log(\sigma)|$ .

*Proof.* Note that this corollary is exactly the same as Theorem 3. In fact, since  $\beta \leq 2/3$ ,

$$\delta_\ell = \min(\sigma^{2/3}|\log^{2/3} \sigma|, \sigma^{1-\beta/2}|\log \sigma|) = \sigma^{1-\beta/2}|\log \sigma|.$$

Therefore,  $\delta_s = \sigma^{1+\beta/2}|\log \sigma|$  and therefore, according to Theorem 3

$$R(H^\alpha(C), \hat{f}^O) = O(\sigma_s \text{Poly}(|\log \sigma|)) = O(\sigma^{1+\beta/2} \text{Poly}(|\log \sigma|)).$$

$\square$

This corollary suggests that, as long as we have an edge orientation estimate that improves as  $\sigma \rightarrow 0$  (or the number of pixels increases), OANLM outperforms NLM. Also note that, as  $\beta$  decreases, the neighborhoods become more isotropic.

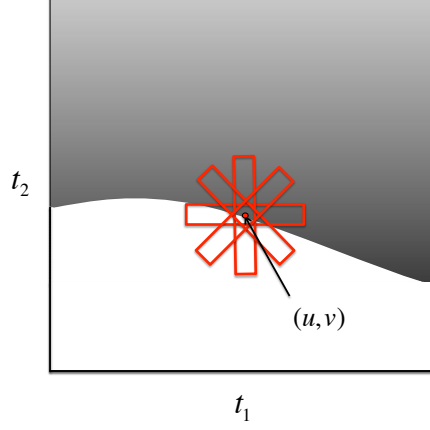


Figure 7: DANLM neighborhoods around one pixel location for  $q = 4$ .

### 3.3. Discrete angle ANLM

In this section, we introduce the Discrete-angle ANLM (DANLM) algorithm that achieves the minimax rate without oracle information. The key idea is to calculate the neighborhood distance over several directional neighborhoods and fuse them to obtain a similarity measure that works well for all directions. As for OANLM, set  $\delta_s = 4\sigma^{4/3}|\log(\sigma)|^{4/3}$  and  $\delta_\ell = 2\sigma^{2/3}|\log(\sigma)|^{2/3}$ , and let  $q \triangleq \pi\sigma^{-2/3}$ . Define the angles  $\theta_0 \triangleq 0$ ,  $\theta_1 \triangleq \sigma^{2/3}$ ,  $\theta_2 \triangleq 2\sigma^{2/3}, \dots, \theta_q \triangleq \pi - \sigma^{2/3}$  and

$$\mathcal{O} \triangleq \{\theta_0, \dots, \theta_q\}.$$

For a point  $(u, v)$  we consider all of the anisotropic, directional neighborhoods for  $\theta \in \mathcal{O}$ . Figure 7 displays four of these neighborhoods.

Define the discrete angle anisotropic distance between  $dY(t_1, t_2)$  and  $dY(s_1, s_2)$  as

$$d_{\mathcal{A}}^2(dY(t_1, t_2), dY(s_1, s_2)) \triangleq \min_{\theta \in \mathcal{O}} d_{\theta, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(s_1, s_2)),$$

where  $d_{\theta, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(s_1, s_2))$  is defined in (6). The DANLM estimate is then given by

$$\hat{f}^D(t_1, t_2) = \frac{\int_{(s_1, s_2) \in S} w_{t_1, t_2}^{AN}(s_1, s_2) X(s_1, s_2) ds_1 ds_2}{\int_{(s_1, s_2) \in S} w_{t_1, t_2}^{AN}(s_1, s_2) ds_1 ds_2}.$$



where

$$w_{t_1, t_2}^{AN}(s_1, s_2) = \begin{cases} 1 & \text{if } d_{\mathcal{A}}^2(dY(t_1, t_2), dY(s_1, s_2)) \leq \frac{2n_s n_\ell \sigma^2}{\delta_s \delta_\ell} + \tau_\sigma, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

In summary, we note the following features of the DANLM algorithm. First, it uses the quadratic scaling  $\delta_s = \delta_\ell^2$ . Second, the optimal neighborhood direction can change from pixel to pixel. The following theorem, proved in Section 7, shows that the risk of this algorithm is within the logarithmic factor of the optimal minimax risk.

**Theorem 4.** *The risk of DANLM satisfies*

$$R(H^\alpha(C_\alpha), \hat{f}^D) \leq O(\sigma^{4/3} |\log(\sigma)|^{4/3}).$$

Figure 8 reprises the simulation experiment of Figure 4 using the four algorithms described thus far: NLM, OANLM with perfect knowledge of the edge orientation, OANLM with imperfect knowledge of the edge orientation, and DANLM with four angles. All three of the latter algorithms outperform the isotropic NLM.

## 4. Practical ANLM algorithms and experiments

In this section we introduce a practical gradient-based ANLM algorithm and then complement the above theoretical arguments with additional simulations and experiments with synthetic and real-world imagery. Note that the algorithms we introduce in this section do not immediately outperform state-of-the-art denoising algorithms such as BM3D [13] but are rather intended to highlight that anisotropic neighborhoods are more suitable for edges than isotropic ones. Since NLM is a key building block in several top-performing algorithms [13, 29], we expect that anisotropy will pay off for such algorithms as it does for NLM. But addressing this important question is left for future research.

### 4.1. Extension to discrete images

In practice the observations are noisy pixelated values of an image and the objective is only to estimate the pixelated values. In this section we



Figure 8: Continuation of the experiment of Figure 4 that compares (a) isotropic NLM, (b) OANLM with perfect knowledge of the edge direction, (c) OANLM with 10% error in the knowledge of the edge direction, (d) DANLM with  $q = 4$  angles. The images are of size  $256 \times 256$  pixels. In the ANLM algorithms,  $\delta_s$  and  $\delta_\ell$  are the same as in Figure 4. The neighborhood size of NLM is  $\sqrt{\delta_s \times \delta_\ell}$ . The edge orientation is  $135^\circ$ . The standard deviation of the noise is  $\xi = 0.5$  which implies PSNR = 6.1dB.

explain how the ideas of directional neighborhood and ANLM can be extended to the discrete settings. Suppose we are interested in estimating a

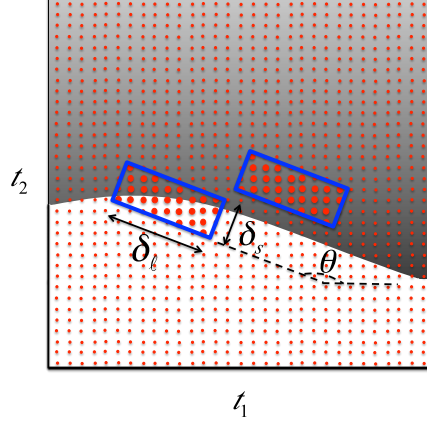


Figure 9: The anisotropic neighborhood  $\bar{I}_{\theta, \delta_s, \delta_\ell}(\frac{i}{n}, \frac{j}{n})$  in the discrete setting for two different pixels of a Horizon class image. The bold red dots represent the anisotropic neighborhoods.

$n \times n$  image  $f(\frac{i}{n}, \frac{j}{n})$  with noisy observations  $o_{i,j} = f(\frac{i}{n}, \frac{j}{n}) + \zeta_{i,j}$ , where  $\zeta_{i,j} \stackrel{iid}{\sim} N(0, \xi^2)$  and  $\xi$  is the standard deviation of the noise. The extension of the anisotropic neighborhood to the discrete setting is straightforward. Let  $\bar{S} \triangleq \{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\} \times \{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$  and  $\tilde{S} \triangleq \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ . For a given set  $B \subset S = [0, 1]^2$  we define  $\bar{B} \triangleq B \cap \bar{S}$ .

The discrete  $(\theta, \delta_s, \delta_\ell)$ -distance between two pixels  $o_{i,j}$  and  $o_{m,\ell}$  is defined as

$$\bar{d}_{\theta, \delta_s, \delta_\ell}^2(o_{i,j}, o_{m,\ell}) \triangleq \frac{1}{|\mathcal{P}|} \sum_{(r,q) \in \mathcal{P}} (o_{i+r, j+q} - o_{m+r, \ell+q})^2, \quad (10)$$

where  $\mathcal{P} \triangleq \{(r, q) \in \mathbb{Z}^2 \mid (\frac{i+r}{n}, \frac{j+q}{n}) \in \bar{I}_{\theta, \delta_s, \delta_\ell}(\frac{i}{n}, \frac{j}{n})\}$ . See Figure 9. Note that for the pixels that are close to the image boundaries the neighborhoods are not rectangular any more and they include the part of the rectangle that is inside the image boundaries. Such non-rectangular neighborhoods will be compared with non-rectangular neighborhoods of the other pixels for the calculation of the distance. The ANLM estimate at pixel  $f(\frac{i}{n}, \frac{j}{n})$  is given by

$$\hat{f}_{i,j}^{\theta, \delta_s, \delta_\ell} \triangleq \frac{\sum_m \sum_\ell \bar{w}_{i,j}^{\theta, \delta_s, \delta_\ell}(m, \ell) o_{m,\ell}}{\sum_m \sum_\ell \bar{w}_{i,j}^{\theta, \delta_s, \delta_\ell}(m, \ell)},$$

where the weights are obtained from the distances  $\bar{d}_{\theta, \delta_s, \delta_\ell}^2(o_{i,j}, o_{m,\ell})$ . For instance, the simple policy introduced in (7) corresponds to

$$\bar{w}_{i,j}^{\theta, \delta_\ell, \delta_s}(m, \ell) \triangleq \begin{cases} 1 & \text{if } \bar{d}_{\theta, \delta_s, \delta_\ell}^2(o_{i,j}, o_{m,\ell}) \leq 2\xi^2 + \tau_n, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Using the discrete anisotropic neighborhood and distance, the extensions of OANLM and DANLM to the discrete setting is straightforward.

#### 4.2. Gradient-based ANLM

While DANLM is somewhat practical and also theoretically optimal for the Horizon class of images, its computational complexity is higher than NLM and grows linearly in the number of directions  $q$ , where  $q \sim o(n^{2/3})$ . As a more practical alternative, we propose Gradient based ANLM (GANLM), which adjusts the orientation of an anisotropic ANLM neighborhood using an estimate of the edge orientation provided by the image gradients. Pseudocode is given in Algorithm 1. Note that GANLM reverts back to NLM in regions with low image gradients, since they will not be “edgy” enough to benefit from the special treatment.

There is a rich literature on robust image gradient estimation [6, 22, 39]. In our implementations we use the simplest method for estimating the image gradient, i.e.,

$$\begin{aligned} g_h(i, j) &= \hat{f}_{i+1,j} - \hat{f}_{i,j}, \\ g_v(i, j) &= \hat{f}_{i,j+1} - \hat{f}_{i,j}, \end{aligned}$$

where  $\hat{f}(i, j)$  is an estimate of the image (for instance from isotropic NLM). If  $g_h(i, j)$  and  $g_v(i, j)$  are the estimated image derivatives at pixel  $(i, j)$ , then we can estimate the local orientation of an edge by  $\hat{\theta}(i, j) = \arctan\left(\frac{g_v(i,j)}{g_h(i,j)}\right)$ . To allay any concerns that gradient-based adaptivity is not robust to noise and errors, we recall Theorem 3, which establishes the robustness of OANLM to edge angle estimation error. For extremely noisy images, numerous heuristics are possible, including estimating the image gradients for GANLM from an isotropic NLM pilot estimate. Figure 11 builds on Figure 8 with a slightly more realistic, curved edge and demonstrates that the pilot estimate approach to GANLM performs almost as well as an oracle GANLM that has access to the edge gradient.

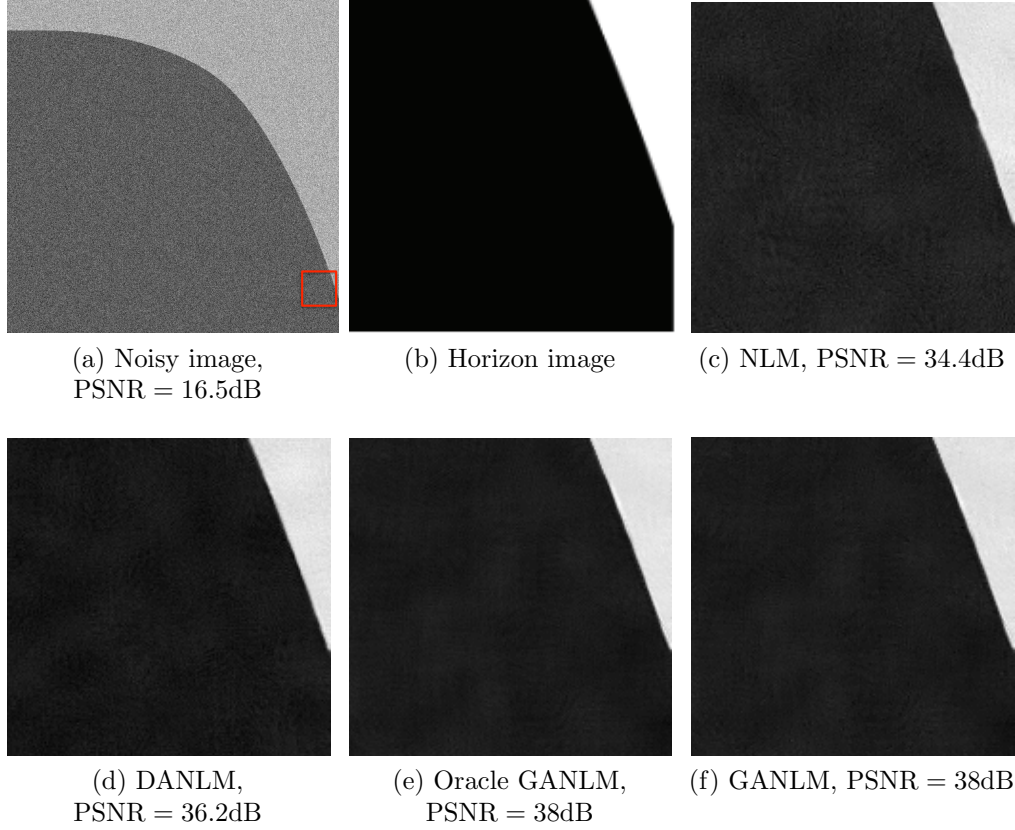


Figure 10: Simulation experiment in the same vein as Figures 4 and 8 but with a slightly more realistic  $C^2$  curved edge. The images are of size  $512 \times 512$  pixels. In the ANLM algorithms  $\delta_s$  and  $\delta_\ell$  follow the quadratic scaling of Theorem 2. The neighborhood size of NLM is  $\sqrt{\delta_s \times \delta_\ell}$ . The standard deviation of the noise is  $\xi = 0.15$ . To show the differences we have zoomed on the area represented by the red block in Figure (a).

---

**Algorithm 1** *Gradient-based ANLM (GANLM)*

---

**Inputs:** $\hat{f}_{i,j}$  : Estimate of the image pixel $\delta_s \times \delta_\ell$ : Size of the neighborhood $\lambda$ : Threshold that determines isotropic/anisotropic neighborhood selectionEstimate image gradient  $(g_h(i, j), g_v(i, j))$  at each pixel  $(i, j)$ **for** every pixel  $(i, j) \in I$  **do**

$$g(i, j) = \sqrt{g_h^2(i, j) + g_v^2(i, j)}$$

$$\theta_{i,j} = \arctan\left(\frac{g_v(i, j)}{g_h(i, j)}\right)$$

**if**  $g_i \geq \lambda$  **then**Perform ANLM at pixel  $y_{i,j}$  with  $d_{\delta_\ell, \delta_s, \theta_{i,j}}$ **else**Perform NLM at pixel  $y_{i,j}$  with  $\sqrt{\delta_s \delta_\ell}$ .**end if****end for**

---

Table 2 summarizes the performance of the algorithms introduced in this paper with that of NLM on the natural test images Barbara [1], Boats [2], and Wet Paint [3] submerged in noise of two different levels. The table enables us to draw two conclusions. First, the performance of the practical GANLM algorithm is very close to the oracle GANLM algorithm. Second these two algorithms outperform DANLM in all but one case and significantly outperform standard NLM in all cases. We use 4 discrete angles<sup>10</sup> in DANLM and attribute the superior performance of GANLM over DANLM to a more accurate selection of orientations.

#### 4.3. Computational complexity of ANLM

Improving the risk performance of linear filters by nonlocal averaging is achieved at the price of higher computational cost. If  $|\mathcal{P}|$  denotes the neighborhood size, then the computational complexity of NLM is proportional to  $\Theta(n^4|\mathcal{P}|)$  as compared to  $\Theta(n^2)$  for linear filters. Such high complexity

---

<sup>10</sup>Experiments with larger  $q$  did not differ appreciably in PSNR at this image size.

Table 2: PSNR Performance of NLM and ANLM algorithms on three natural test images at two noise levels.

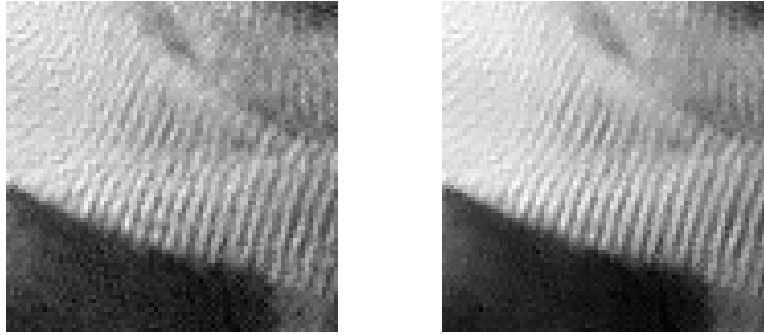
Test image	Algorithm	$\xi = 0.25, 12.1\text{dB}$	$\xi = 0.15, 16.5\text{dB}$
Barbara (512 x 512)	NLM	22.48	25.86
	DANLM	23.15	26.27
	Oracle GANLM	23.51	26.63
	GANLM	23.50	26.60
Boats (512 x 512)	NLM	22.75	25.83
	DANLM	23.45	26.45
	Oracle GANLM	23.88	26.45
	GANLM	23.75	26.35
Wet Paint (1024 x 1024)	NLM	27.66	30.49
	DANLM	28.41	30.75
	Oracle GANLM	29.02	31.18
	GANLM	28.86	31.07

is not acceptable in many applications. Therefore, several effective methods have been proposed to reduce the computational complexity of NLM. See [7, 8, 40, 45] and the references therein for more information on the research in this direction. These approaches have reduced the computational complexity of NLM to an acceptable limit for practice.

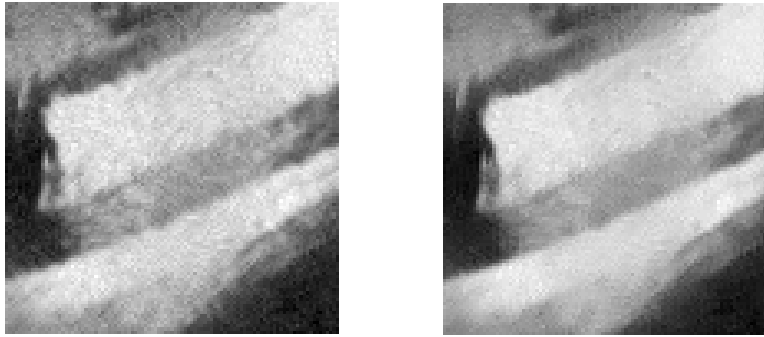
The computational complexity of ANLM is in general higher than that of NLM. For instance, the computational complexity of the Gradient-based ANLM algorithm introduced in Section 4.2 is roughly two times that of NLM if we ignore the gradient calculation, which is  $\Theta(n^2)$ . The computational complexity of DANLM is  $\Theta(qn^4|\mathcal{P}|)$ , where  $q$  denotes the number of angles considered. Most of the approaches proposed for speeding up NLM are applicable to ANLM as well. Furthermore, due to specific structure of DANLM, in particular the overlap of the neighborhoods of each pixel, we expect to be able to reduce its computational complexity even further. But this question is left for future research.

## 5. Related work on anisotropic denoising

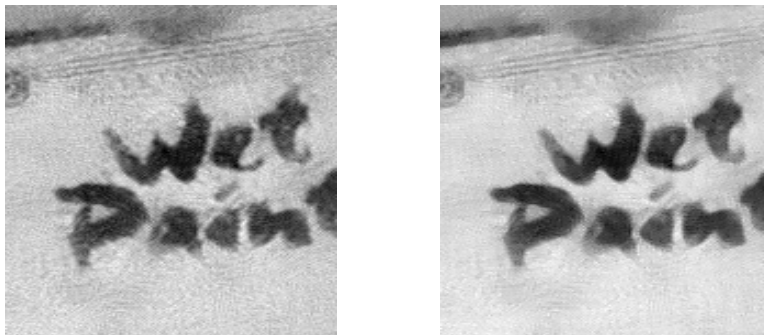
Anisotropy is a fundamental feature of real-world images. As a result, anisotropic image processing tools can be traced back at least as far as the 1970s [44]. Here, we briefly compare and contrast some of the relevant



(a) Barbara



(b) Boats



(c) Wet Paint

Figure 11: Zooms of the natural test images denoised by NLM (left) and ANLM (right) studies in Table 2. The standard deviation of the noise is  $\xi = 0.15$  that implies PSNR= 16.5dB.



anisotropic denoising schemes with ANLM.

*Anisotropic filtering* methods use a space-varying linear convolution filters to reduce blur along image edges. Toward this goal, [44] considers several different neighborhoods around a pixel and selects the most “homogeneous” neighborhood to smooth over. A more advanced version of this idea can be found in [20, 55, 56]. There are major differences between such algorithms and NLM/ANLM; in particular, the estimators are local and do not exploit global similarities. See also [43] for a more complete overview of such algorithms and their differences with NLM/ANLM.

*Anisotropic diffusion* methods smooth a noisy image with a Gaussian kernel. As the standard deviation of the kernel increases, the smoothing process introduces larger bias to the edges. In [26, 33] the authors proved that the set of images derived by this approach can be viewed as the solution to the heat diffusion equation. Perona and Malik [47] noted the isotropy in the heat equation and introduced anisotropy. Their anisotropic diffusion starts with the heat equation but at every iteration exploits the gradient information of the previous estimate to increase the conductance along the edges and decrease it across the edges. See [30, 31, 58] for more recent advances in this direction. Efforts to theoretically analyze the risk of such algorithms have left many open questions remaining [4]. It is conjectured that there is a connection between such algorithms and iteratively applying nonlinear filters such as the median filter. It is worth noting that the idea of applying an image denoising algorithm iteratively and guiding it at every iteration based on previous estimates goes back to Tukey’s twicing [59].

*Anisotropic transformations* enable simple threshold based denoising algorithms. While the standard separable wavelet transform cannot exploit the smoothness of the edge contour, a menagerie of anisotropic transforms have been developed, including ridgelets [11], curvelets [9], wedgelets [17], platelets [63, 64], shearlets [19, 25, 35], contourlets [16], bandelets [46, 48], and directionlets [61]. As mentioned in the Introduction, among the above algorithms only wedgelets can obtain the optimal minimax risk for the Horizon class; however wedgelets are not suited to denoising textures. One promising avenue combining wavelets (for texture denoising) and wedgelets (for edge denoising) could follow the path of the image coder in [62]. See [27, 36] for an overview of anisotropic transformations and their applications in image processing.

Alternative NLM algorithms have been proposed to address the inefficiency of using a fixed neighborhood [13–15, 24, 29, 49, 52, 65]. In [15, 52], the authors aggregate several nonlocal estimates to obtain a final estimate. The nonlocal estimates are derived from NLM with different neighborhoods around each pixel. While these methods improve upon NLM on edges in practice, there is no theoretical result to support such empirical results. Similar approaches have been exploited in several state-of-the-art denoising algorithms [13, 14]. In [29, 60, 65], the authors adapt the neighborhood size to the local image content. [29] considers different isotropic NLM neighborhood sizes depending on how smoothly the image content varies. [65] uses image gradient information to increase the weights of the NLM along edges and decrease them across edges. This method is equivalent to modifying the threshold parameter  $t_n$  to force NLM to assign higher weights to edge-like neighborhoods. [60] sets the neighborhood size based on Stein’s unbiased estimate of the risk (SURE). Other generalizations include the rotationally invariant similarity measure [24] and nonlocal variational approaches [23, 32, 49]. Unfortunately, these techniques do not reduce the bias that renders NLM sub-optimal.

Finally, *data-driven optimality criteria* have been considered in [12, 38, 50], where the authors derive lower bounds for the performance of denoising algorithms. However, the analyses provided in these papers are not fully rigorous and do not cover the performance of NLM for images with sharp edges.

## 6. Discussion and future directions

We have introduced and analyzed a framework for anisotropic nonlocal means (ANLM) denoising. Similar to NLM, ANLM exploits nonlocal information in estimating the pixel values. However, unlike NLM, ANLM uses anisotropic, oriented neighborhoods that can exploit the smoothness of edge contours. This enables ANLM to outperform NLM both theoretically and empirically. In fact, the performance of ANLM is within a logarithmic factor of optimal as measured by the minimax rate on the Horizon class of images.

Numerous questions remain open for future research. From the theoretical perspective, the risk analysis of GANLM, the application to noise models beyond Gaussian, and the extension to three dimensions and beyond (for

seismic, medical, and other data) pose interesting research challenges. From a practical perspective, the question of how to best tune ANLM to match the nuanced edges and textures of real-world images remains open, since we have considered only brutal binary images here. Finally, while NLM is no longer the state-of-the-art denoising algorithm, it is a key building block in several top-performing algorithms. It would be interesting to see whether anisotropy pays off as handily for those algorithms as it does for NLM.

## 7. Proofs of the main results

### 7.1. Preamble

We first introduce some notation. Define the following partitions of the set  $S = [0, 1]^2$ :

$$\begin{aligned} S_1 &\triangleq \{(v, u) \mid u > h(v) + (1 + C/2)\delta_s\}, \\ S_2 &\triangleq \{(v, u) \mid h(v) - (1 + C/2)\delta_s < u < h(v) + (1 + C/2)\delta_s\}, \\ S_3 &\triangleq \{(v, u) \mid u < h(v) - (1 + C/2)\delta_s\}. \end{aligned}$$

It is important to note that, if  $(t_1, t_2) \in S_1$  and  $\tan(\theta) = h'(t_1)$ , then  $I_{\theta, \delta_s, \delta_\ell}(t_1, t_2)$  does not overlap with the edge contour. In other words, the correctly aligned neighborhood of  $(t_1, t_2) \in S_1$  is always above the edge. The points in  $S_3$  satisfy a similar property. This is clarified in Figure 12.

We further partition  $S_1$  into  $P_1$  and  $P_2$  and  $S_3$  into  $P_3$  and  $P_4$  such that

$$\begin{aligned} P_1 &\triangleq \{(v, u) \mid h(v) + 2\delta_\ell + (C/2)\delta_s \leq u\}, \\ P_2 &\triangleq \{(v, u) \mid h(v) + (1 + C/2)\delta_s \leq u \leq h(v) + 2\delta_\ell + (C/2)\delta_s\}, \\ P_3 &\triangleq \{(v, u) \mid h(v) - (1 + C/2)\delta_s \geq u \geq h(v) - 2\delta_\ell - (C/2)\delta_s\}, \\ P_4 &\triangleq \{(v, u) \mid u \leq h(v) - 2\delta_\ell - (C/2)\delta_s\}. \end{aligned}$$

**Lemma 1.** *Any neighborhood of pixel  $(v, u) \in P_1$  will lie completely above the edge contour.*

*Proof.* Let  $(t_1, t_2) \in P_1$ . If  $(u, v) \in I_{\theta, \delta_s, \delta_\ell}(t_1, t_2)$ , then

$$t_2 - v < \delta_\ell. \tag{12}$$

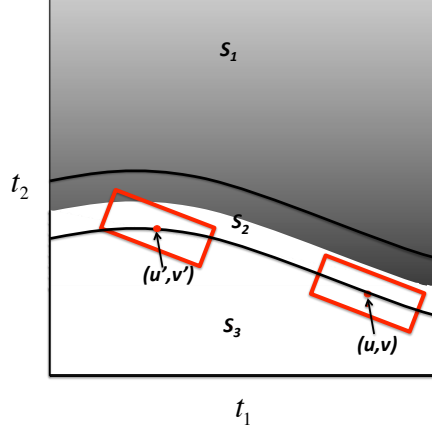


Figure 12: Regions  $S_1$ ,  $S_2$  and  $S_3$ . The neighborhood of  $(u, v) \in S_3$  is aligned with the edge contour, and therefore it does not intersect the edge, while the neighborhood of  $(u', v')$  is not aligned and therefore may intersect with the edge. The neighborhoods of the pixels in  $S_2$  may intersect the edge even though they are correctly aligned.

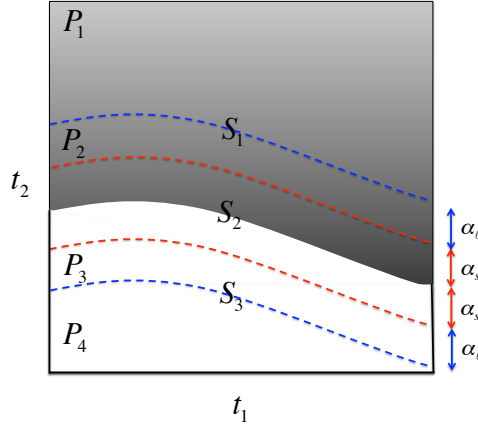


Figure 13: The regions  $P_1$ ,  $P_2$ ,  $P_3$ ,  $P_4$  with  $\alpha_s = (1 + C/2)\delta_s$  and  $\alpha_\ell = 2\delta_\ell - \delta_s$ . Every neighborhood of  $(t_1, t_2) \in P_1$  will lie completely above the edge contour. However, some of the neighborhoods of the pixels  $(t_1, t_2) \in P_2$  may intersect the edge. A similar property holds for the regions  $P_3$  and  $P_4$ .

On the other hand, for  $t'_1 \in [t_1 - \delta_\ell, t_1 + \delta_\ell]$  we have,

$$h(t'_1) = h(t_1) + h'(t_1)(t'_1 - t_1) + \frac{1}{2}h''(t''_1)(t'_1 - t_1)^2,$$

where  $t''_1$  is between  $t_1$  and  $t'_1$ . According to the definition of the Horizon class,  $h \in \text{H\"older}^\alpha(C)$  and hence  $|h'(t_1)| \leq 1$ . Therefore,

$$h(t'_1) - h(t_1) < \delta_\ell + \frac{1}{2}C\delta_s, \quad (13)$$

Comparing (12) and (13) completes the proof.  $\square$

In spite of  $P_1$ , some of the neighborhoods of the pixels  $(v, u) \in P_2$  may intersect the edge. Similarly one can prove that any neighborhood of a pixel  $(v, u) \in P_4$  lies completely below the edge, and some of the neighborhoods of the pixel  $(v, u) \in P_3$  may intersect the edge. Figure 13 displays these regions. The following lemma, proved in [41], will be used in the proofs of the main theorems. For the sake of completeness, we sketch the proof here.

**Lemma 2.** *Let  $Z_1, Z_2, \dots, Z_r$  be iid  $N(0, 1)$  random variables. The  $\chi_r^2$  random variable  $\sum_{i=1}^r Z_i^2$  concentrates around its mean with high probability, i.e.,*

$$\mathbb{P} \left( \frac{1}{r} \sum_i Z_i^2 - 1 > t \right) \leq e^{-\frac{r}{2}(t - \ln(1+t))}, \quad (14)$$

$$\mathbb{P} \left( \frac{1}{r} \sum_i Z_i^2 - 1 < -t \right) \leq e^{-\frac{r}{2}(t + \ln(1-t))}. \quad (15)$$

*Proof.* Here we prove (14); the proof of (15) follows along very similar lines. From Markov's Inequality, we have

$$\begin{aligned} \mathbb{P} \left( \left( \frac{1}{r} \sum_{i=1}^r Z_i^2 \right) - 1 > t \right) &\leq e^{-\eta t - \eta} \mathbb{E} \left( e^{\frac{\eta}{r} \sum_{i=1}^r Z_i^2} \right) \\ &= e^{-\eta t - \eta} \left( \mathbb{E} \left( e^{\frac{\eta Z_1^2}{r}} \right) \right)^r = \frac{e^{-\eta t - \eta}}{\left( 1 - \frac{2\eta}{r} \right)^{\frac{r}{2}}}. \end{aligned} \quad (16)$$

The last inequality follows from Lemma 3 in [41]. The upper bound proved above holds for any  $\eta < \frac{r}{2}$ . To obtain the lowest upper bound we minimize  $\frac{e^{-\eta t - \eta}}{(1 - \frac{2\eta}{r})^{\frac{r}{2}}}$  over  $\eta$ . The optimal value is  $\eta^* = \arg \min_{\eta} \frac{e^{-\eta t - \eta}}{(1 - \frac{2\eta}{r})^{\frac{r}{2}}} = \frac{rt}{2(t+1)}$ . Plugging  $\eta^*$  into (16) proves the lemma.  $\square$

## 7.2. Proof of Theorem 2

In the set up above, we consider three different regions for the point  $(t_1, t_2)$ . As we will see in the proof, the risk of all pixels in each region has the same upper bound. We calculate these upper bounds and then combine them to obtain a master upper bound for the risk of OANLM.

Case I:  $(t_1, t_2) \in S_1$ . We know that if the anisotropic neighborhood of  $(t_1, t_2)$ ,  $I_{\theta, \delta_s, \delta_\ell}$  is aligned with the edge contour, i.e.,  $\tan(\theta) = h'(t_1)$ , then it does not intersect the edge contour. To calculate the OANLM estimate we first calculate the weights. Define

$$\mathbf{z}_{t_1, t_2}^{\theta, \delta_s, \delta_\ell}(j_1, j_2) \triangleq \frac{n_s n_\ell}{\delta_s \delta_\ell} \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_\ell}^{j_1, j_2}} dW(s_1, s_2), \quad (17)$$

and

$$\begin{aligned} Z(t_1, t_2) &\triangleq \frac{n_s n_\ell}{\delta_s \delta_\ell} \int \int_{(s_1, s_2) \in I_{\theta, \frac{\delta_s}{n_s}, \frac{\delta_\ell}{n_\ell}}(t_1, t_2)} dW(s_1, s_2), \\ F(t_1, t_2) &\triangleq \frac{n_s n_\ell}{\delta_s \delta_\ell} \int \int_{(s_1, s_2) \in I_{\theta, \frac{\delta_s}{n_s}, \frac{\delta_\ell}{n_\ell}}(t_1, t_2)} f(s_1, s_2) ds_1 ds_2. \end{aligned}$$

For notational simplicity we will use  $w(s_1, s_2)$  and  $\mathbf{z}_{t_1, t_2}(j_1, j_2)$  instead of  $w_{t_1, t_2}^{\theta, \delta_s, \delta_\ell}(s_1, s_2)$  and  $\mathbf{z}_{t_1, t_2}^{\theta, \delta_s, \delta_\ell}(j_1, j_2)$ , respectively. Define

$$\begin{aligned} A_1 &\triangleq \{(u, v) \in P_1 \mid w(u, v) = 1\}, \\ A_2 &\triangleq \{(u, v) \in P_4 \mid w(u, v) = 0\}. \end{aligned}$$

Here, let  $\lambda(\cdot)$  denote the Lebesgue measure of a set over  $\mathbb{R}^2$ .

**Lemma 3.** Let  $\delta_\ell = 2\sigma^{2/3}|\log(\sigma)|^{2/3}$  and  $\delta_s = 4\sigma^{4/3}|\log(\sigma)|^{4/3}$ ,  $n_s = 2|\log(\sigma)|^{2/3}$ ,  $n_\ell = 4|\log(\sigma)|^{4/3}$ , and  $\tau_\sigma = \frac{2}{\sqrt{|\log \sigma|}}$ . Then,

$$\begin{aligned}\mathbb{P}(\lambda(P_1) - \lambda(A_1) > \epsilon) &= O\left(\frac{\sigma^8}{\epsilon}\right), \\ \mathbb{P}(\lambda(P_4) - \lambda(A_2) > \epsilon) &= O\left(\frac{\sigma^8}{\epsilon}\right).\end{aligned}$$

*Proof.* Here we prove the first expression. The proof of the second expression follows along a similar route. Let  $(u, v) \in P_1$ . Since  $\mathbf{z}_{t_1, t_2}(j_1, j_2)$  is the integral of the Wiener sheet, we have  $\mathbf{z}_{t_1, t_2}(j_1, j_2) \sim N(0, \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell})$ . Similarly,  $\mathbf{z}_{u, v}(j_1, j_2) \sim N(0, \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell})$ . Combining this with Lemma 2 we conclude that

$$\mathbb{P}\left(d_{\theta, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - 2\frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell} \geq \tau_\sigma\right) \leq e^{-\frac{n_s n_\ell t^2}{4}} = O(\sigma^8).$$

Therefore,

$$\begin{aligned}\mathbb{E}(\lambda(A_1)) &= \mathbb{E} \int_{(u, v) \in P_1} I((u, v) \in A_1) = \int_{(u, v) \in P_1} \mathbb{P}((u, v) \in A_1) \\ &\geq \int_{(u, v) \in P_1} \mathbb{P}\left(d_{\theta, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - 2\frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell} \leq \tau_\sigma\right) \\ &= \lambda(P_1) - O(\sigma^8).\end{aligned}\tag{18}$$

An upper bound for  $\mathbb{P}(\lambda(P_1) - \lambda(A_1) > \epsilon)$  using the Markov inequality yields the result

$$\mathbb{P}(\lambda(P_1) - \lambda(A_1) > \epsilon) \leq \frac{\mathbb{E}(\lambda(P_1) - \lambda(A_1))}{\epsilon} = O\left(\frac{\sigma^8}{\epsilon}\right).$$

□

This lemma indicates that most of the points in  $P_1$  may contribute in the estimation of  $(t_1, t_2) \in S_1$  and most of the points in  $P_4$  do not contribute in the estimate. Define the event  $\mathcal{E}$  as

$$\mathcal{E} \triangleq \{\lambda(P_1) - \lambda(A_1) < \sigma^2\} \cap \{\lambda(P_4) - \lambda(A_2) < \sigma^2\}.$$

Note that according to Lemma 3,  $\mathbb{P}(\mathcal{E}^c) = O(\sigma^6)$ . As we will see later this lemma simplifies the calculation of the risk dramatically. For notational simplicity we also define the following notations:

$$\begin{aligned} wXdudv &\triangleq w(u, v)X(u, v)dudv, \\ Xdudv &\triangleq X(u, v)dudv, \\ wdudv &\triangleq w(u, v)dudv, \\ Zdudv &\triangleq Z(u, v)dudv, \\ wZdudv &\triangleq w(u, v)Z(u, v)dudv. \end{aligned}$$

The risk of the OANLM estimator at  $(t_1, t_2) \in S_1$  is then given by

$$\mathbb{E} \left( \frac{\int_S wXdudv}{\int_S wdudv} \right)^2.$$

Define  $U = \left( \frac{\int_S wXdudv}{\int_S wdudv} \right)^2$ . We have

$$\mathbb{E}(U) = \mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{E}(U \mid \mathcal{E}^c)\mathbb{P}(\mathcal{E}^c) \leq \mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}^c). \quad (19)$$

Lemma 3 proves that  $\mathbb{P}(\mathcal{E}^c) = O(\sigma^6)$ . Also, since for  $(u, v) \in A_1$   $w(u, v) = 1$ , we have

$$\begin{aligned} &\left| \int_{P_1} wXdudv - \int_{P_1} Xdudv \right| \\ &= \left| \int_{P_1} wXdudv - \int_{A_1} wXdudv + \int_{A_1} Xdudv - \int_{P_1} Xdudv \right| \\ &= \left| \int_{P_1} wXdudv - \int_{A_1} wXdudv \right| + \left| \int_{A_1} Xdudv - \int_{P_1} Xdudv \right| \\ &= O(\lambda(P_1 \setminus A_1)). \end{aligned} \quad (20)$$

For the last inequality we have assumed that the estimate is bounded over the  $P_1$  region. This assumption can be also justified by calculating the probability that this condition does not hold and showing that the probability is negligible. Using arguments similar to (20) it is straightforward to prove that



$$\begin{aligned}
\left| \int_{P_4} wX dudv \right| &= O(\lambda(P_4 \setminus A_2)). \\
\left| \int_{P_1} wdudv - \int_{P_1} dudv \right| &= O(\lambda(P_1 \setminus A_1)). \\
\left| \int_{P_4} wdudv \right| &= O(\lambda(P_4 \setminus A_2)).
\end{aligned} \tag{21}$$

Define  $P_{14} \triangleq P_1 \cup P_4$  and  $B \triangleq S \setminus P_{14}$ . We now calculate an upper bound for the first term of (33)

$$\begin{aligned}
&\mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) \\
&\stackrel{(a)}{=} \mathbb{E} \left( \left( \frac{\int_{P_{14}} wX dudv + \int_B wX dudv}{\int_{P_{14}} wdudv + \int_B wdudv} \right)^2 \mid \mathcal{E} \right) \mathbb{P}(\mathcal{E}) \\
&= \mathbb{E} \left( \left( \frac{\int_{P_1} Xdudv + \int_B wX dudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 \mid \mathcal{E} \right) \mathbb{P}(\mathcal{E}) + O(\sigma^2) \\
&\leq \mathbb{E} \left( \frac{\int_{P_1} Xdudv + \int_B wX dudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 + O(\sigma^2) \\
&\stackrel{(b)}{\leq} \mathbb{E} \left( \frac{\int_B wFdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 + \mathbb{E} \left( \frac{\int_{P_1} Zdudv + \int_B wZdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 \\
&\quad + 2 \sqrt{\mathbb{E} \left( \frac{\int_B wFdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2} \sqrt{\mathbb{E} \left( \frac{\int_{P_1} Zdudv + \int_B wZdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2} \\
&\quad + O(\sigma^2).
\end{aligned} \tag{22}$$

Equality (a) is an application of (20) and (21). To obtain (b) we plug in the value of  $X = F + Z$ , where  $F$  and  $Z$  represent signal and noise respectively. In the next two lemmas, Lemma 22 and Lemma 5, we obtain upper bounds for individual terms in (22).

**Lemma 4.** *Let  $w(u, v)$  be the weights of OANLM with  $\delta_\ell, \delta_s, n_s, n_\ell$ , and  $\tau_\sigma$*

as specified in Lemma 3. Then

$$\mathbb{E} \left( \frac{\int_B w F dudv}{\int_{P_1} dudv + \int_B w dudv} \right)^2 \leq O(\sigma^{4/3} |\log(\sigma)|^{4/3}).$$

*Proof.* We have

$$\begin{aligned} \mathbb{E} \left( \frac{\int_B w F dudv}{\int_{P_1} dudv + \int_B w dudv} \right)^2 &\stackrel{(a)}{\leq} \mathbb{E} \left( \frac{\int_B dudv}{\int_{P_1} dudv} \right)^2 \\ &= \left( \frac{\lambda(B)}{\lambda(P_1)} \right)^2 = O(\sigma^{4/3} |\log(\sigma)|^{4/3}). \end{aligned}$$

To obtain Inequality (a) we maximize the numerator and minimize the denominator independently. The last equality is due to the fact that  $\lambda(B) = O(\delta_\ell)$ .  $\square$

**Lemma 5.** Let  $w(u, v)$  be the weights of OANLM with  $\delta_\ell, \delta_s, n_s, n_\ell$ , and  $\tau_\sigma$  as specified in Lemma 3. Then

$$\mathbb{E} \left( \frac{\int_{P_1} Z dudv + \int_B w Z dudv}{\int_{P_1} dudv + \int_B w dudv} \right)^2 \leq O(\sigma^{4/3} |\log(\sigma)|^{4/3}).$$

*Proof.* Since  $\int_B w dudv \geq 0$ , and we are interested in the upper bound of the risk, we can remove it from the denominator to obtain

$$\begin{aligned} \mathbb{E} \left( \frac{\int_{P_1} Z dudv + \int_B w Z dudv}{\int_{P_1} dudv + \int_B w dudv} \right)^2 &\leq \mathbb{E} \left( \frac{\int_{P_1} Z dudv + \int_B w Z dudv}{\int_{P_1} dudv} \right)^2 \\ &\leq \underbrace{\mathbb{E} \left( \frac{\int_{P_1} Z dudv}{\int_{P_1} dudv} \right)^2}_{V_1} + \underbrace{\mathbb{E} \left( \frac{\int_B w Z dudv}{\int_{P_1} dudv} \right)^2}_{V_2} \\ &\quad + \underbrace{2 \sqrt{\mathbb{E} \left( \frac{\int_{P_1} Z dudv}{\int_{P_1} dudv} \right)^2} \sqrt{\mathbb{E} \left( \frac{\int_B w Z dudv}{\left( \int_{P_1} dudv \right)^2} \right)^2}}_{V_3}. \end{aligned}$$

Now we obtain upper bounds for  $V_1, V_2$ , and  $V_3$  separately. First, we have

$$\begin{aligned}
V_1 &= \mathbb{E} \left( \frac{\int_{P_1} Z du dv}{\int_{P_1} du dv} \right)^2 = \mathbb{E} \left( \frac{\int_{P_1} \int_{P_1} Z(u, v) Z(u', v') du' dv' du dv}{\left( \int_{P_1} du dv \right)^2} \right) \\
&\stackrel{(b)}{=} \left( \frac{\int_{P_1} \int_{I_{\theta, \frac{2\delta_s}{n_s}, \frac{2\delta_\ell}{n_\ell}}(u, v)} \mathbb{E}(Z(u, v) Z(u', v')) du' dv' du dv}{\left( \int_{P_1} du dv \right)^2} \right) \\
&\stackrel{(c)}{\leq} \left( \frac{\int_{P_1} \int_{I_{\theta, \frac{2\delta_s}{n_s}, \frac{2\delta_\ell}{n_\ell}}(u, v)} \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell} du' dv' du dv}{\left( \int_{P_1} du dv \right)^2} \right) = O(\sigma^2). \tag{23}
\end{aligned}$$

Equality (b) is due to the fact that if  $(u', v') \notin I_{\theta, \frac{2\delta_s}{n_s}, \frac{2\delta_\ell}{n_\ell}}(u, v)$  then  $Z(u, v)$  and  $Z(u', v')$  will be independent and hence  $\mathbb{E}(Z(u, v) Z(u', v')) = 0$ . Inequality (c) is due to the Cauchy-Schwartz inequality  $|\mathbb{E}(Z(u, v) Z(u', v'))| \leq \sqrt{\mathbb{E}Z^2(u, v)} \sqrt{\mathbb{E}Z^2(u', v')}$ .

To obtain an upper bound for  $V_2$ , we first note that

$$\begin{aligned}
&\mathbb{E}(w(u, v) Z(u, v) w(u', v') Z(u', v')) \\
&\leq \sqrt{\mathbb{E}(w(u, v) Z(u, v))^2} \sqrt{\mathbb{E}(w(u', v') Z(u', v'))^2} = \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell}, \tag{24}
\end{aligned}$$

and hence,

$$\begin{aligned}
V_2 &= \mathbb{E} \left( \frac{\int_B w Z du dv}{\int_{P_1} du dv} \right)^2 \\
&= \mathbb{E} \left( \frac{\int_B \int_B w(u, v) Z(u, v) w(u', v') Z(u', v') du dv du' dv'}{\left( \int_{P_1} du dv \right)^2} \right) \\
&\leq \left( \frac{\int_B \int_B \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell} du dv du' dv'}{\left( \int_{P_1} du dv \right)^2} \right) = O(\delta_\ell^2). \tag{25}
\end{aligned}$$

Note that the last inequality is due to the fact that  $\lambda(B) = O(\delta_\ell)$ . Using (23) and (25) it is straightforward to obtain an upper bound for  $V_3$  by using the

Cauchy-Schwartz inequality. Combining the upper bounds for  $V_1$ ,  $V_2$ , and  $V_3$  completes the proof.  $\square$

Combining (22) with Lemmas 4 and 5 establishes the following upper bound for Case I, where  $(t_1, t_2) \in S_1$ :

$$\mathbb{E} \left( f(t_1, t_2) - \frac{\int_S w(u, v) X(u, v) dudv}{\int_S w(u, v) dudv} \right)^2 = O(\sigma^{4/3} |\log(\sigma)|^{4/3}). \quad (26)$$

Case II:  $(t_1, t_2) \in S_2$ . In this case we assume that the risk is bounded by 1, since the function  $f$  is bounded between 0 and 1. If the estimate is out of this range, then we will map it to either 0 or 1.

Case III:  $(t_1, t_2) \in S_3$ . This case is exactly the same as Case I, and hence we skip the proof.

Finally, combining our results for Cases I, II, and III, we can calculate an upper bound for the risk of OANLM as

$$\begin{aligned} R(f, \hat{f}^O) &= \mathbb{E}(\|f - \hat{f}^O\|^2) = \int_{S_1 \cup S_3} (f - \hat{f}^O)^2 dt_1 dt_2 + \int_{S_2} (f - \hat{f}^O)^2 dt_1 dt_2 \\ &= \lambda(S_1 \cup S_3) O(\sigma^{4/3} |\log(\sigma)|^{4/3}) + \lambda(S_2) O(1) = O(\sigma^{4/3} |\log(\sigma)|^{4/3}). \end{aligned}$$

This ends the proof of Theorem 2.

### 7.3. Proof of Corollary 1

The proof of this corollary follows exactly the same route as that of Theorem 2. We merely redefine the regions  $S_k$  and  $P_k$  for  $k \in \{1, 2, 3\}$ . The new definition of the  $S_k$  regions for the NLM algorithm is given by

$$\begin{aligned} S_1^n &\triangleq \{(t_1, t_2) \mid t_2 > h(t_1) + 2\delta\}, \\ S_3^n &\triangleq \{(t_1, t_2) \mid t_2 < h(t_1) - 2\delta\}, \end{aligned}$$

and  $S_2^n \triangleq S \setminus S_1^n \cup S_3^n$ . Since the neighborhoods in NLM are isotropic, we require a single parameter to describe the neighborhood size,  $\delta = \delta_s = \delta_\ell$ . For notational simplicity we have assumed that  $\frac{C}{2}\delta^2 < \delta$ . This is a valid assumption for small enough values of  $\sigma$ , since  $\delta \rightarrow 0$  as  $\sigma \rightarrow 0$ . This assumption implies that the neighborhood of the pixels in  $S_1^n$  and  $S_3^n$  does

not intersect with the edge contour. Since the neighborhoods no longer have different anisotropic lengths, the size of the intersection of the neighborhoods and the edge contour is identical in all directions and there is no need to define the regions,  $P_1, \dots, P_4$ . Equivalently, we set  $P_1 = S_1^n$ ,  $P_4 = S_3^n$  and  $P_2 = P_3 = \emptyset$ . With these definitions the rest of the proof follows along the exact same lines as the proof of Theorem 2. In particular one can show that the risk of a pixel  $(t_1, t_2) \in S_1^n \cup S_3^n$  is  $O(\sigma^2 \log^2(\sigma))$ , and the risk of a point  $(t_1, t_2) \in S_2^n$  is  $O(1)$ . Furthermore,  $\lambda(S_2) = O(2\delta) = O(\sigma \log(\sigma))$ . Therefore the entire risk of NLM is equal to

$$\begin{aligned} R(f, \hat{f}^N) &= \int_{S_1 \cup S_3} (f - \hat{f}^N)^2 dt_1 dt_2 + \int_{S_2} (f - \hat{f}^N)^2 dt_1 dt_2 \\ &= \lambda(S_1 \cup S_3) O(\sigma^2 \log^2(\sigma)) + \lambda(S_2) O(1) = O(\sigma |\log(\sigma)|). \end{aligned}$$

#### 7.4. Proof of Theorem 3

The proof of this theorem is very similar to the proof of Theorem 2. The only difference is in the definitions of  $S_k$  and  $P_k$  for  $k \in \{1, 2, 3\}$ . Since there is a mismatch between the orientations of the neighborhood and the edge contour, the neighborhood of a point in  $S_1$  may intersect the edge contour. In order to fix this, we define the new regions called  $S_i^\beta$  and  $P_i^\beta$ . If the error in  $\theta$  is upper bounded by  $c_\beta \sigma^\beta$ , then define

$$\begin{aligned} S_1^\beta &\triangleq \{(t_1, t_2) \mid t_2 > h(t_1) + c_\beta \sigma^\beta \delta_\ell + \delta_s + (C/2) \delta_\ell^2\}, \\ S_3^\beta &\triangleq \{(t_1, t_2) \mid t_2 < h(t_1) - c_\beta \sigma^\beta \delta_\ell - \delta_s - (C/2) \delta_\ell^2\}, \end{aligned}$$

and  $S_2^\beta \triangleq S \setminus S_1^\beta \cup S_3^\beta$ . Let  $(t_1, t_2) \in S_1^\beta$  and set  $\theta = \arctan(h'(t_1))$ . Consider a directional neighborhood of  $(t_1, t_2)$  with direction  $\hat{\theta}$ , where  $|\hat{\theta} - \theta| \leq c_\beta \sigma^\beta$ . Then it is straightforward to confirm that

$$I_{\hat{\theta}, \delta_s, \delta_\ell}(t_1, t_2) \subset S_1^\beta.$$

Furthermore, define  $P_1^\beta \triangleq P_1$ ,  $P_4^\beta \triangleq P_4$ ,  $P_2^\beta \triangleq S_1^\beta \setminus P_1$ , and  $P_3^\beta \triangleq S_3^\beta \setminus P_4$ . Note that the risk of a point  $(t_1, t_2) \in S_1^\beta$  can be exactly calculated as in the proof of Theorem 2 and therefore, it is upper bounded by  $O(\sigma^{4/3} |\log(\sigma)|^{4/3})$ . Also, the risk of the pixels in  $S_2^\beta$  upper bounded by  $O(1)$ . Finally the Lebesgue measure of  $S_2^\beta$  is  $O(c_\beta \sigma^\beta \delta_\ell + \delta_s + (C/2) \delta_\ell^2)$ . Combining these facts establishes the claimed upper bound.

### 7.5. Proof of Theorem 4

Since the proof is mostly similar to that of Theorem 2, we shall focus on the major differences. The first difference is that we consider the regions  $P_1$ – $P_4$  instead of  $S_1$ – $S_3$ . Previously the regions  $P_1$  and  $P_2$  were treated jointly under the region  $S_1$ . Instead, we shall now consider  $P_1$  and  $P_2$  separately, and their differences shall become progressively apparent.

Case I:  $(t_1, t_2) \in P_1$ . We start with the calculation of the weights  $w(u, v)$  for  $(u, v) \in P_1 \cup P_4$ . Define

$$\begin{aligned} R_1 &\triangleq \{(u, v) \in P_1 \mid w^D(u, v) = 1\}, \\ R_2 &\triangleq \{(u, v) \in P_4 \mid w^D(u, v) = 0\}. \end{aligned}$$

Intuitively speaking, we expect most of the pixels in  $P_1$  to contribute to the estimate of the pixel at  $(t_1, t_2)$ , since all their directional neighborhoods are very similar to the directional neighborhoods of  $(t_1, t_2)$ . This implies that  $R_1$  is “close” to  $P_1$ . Also, we expect that most of the pixels in  $P_4$  do not contribute in the estimate since all their directional neighborhoods are different from those of  $(t_1, t_2) \in P_1$ . Lemma 6 provides a formal statement of this intuition.

**Lemma 6.** *Let  $\delta_s, \delta_\ell, n_s, n_\ell, t$  be as defined in Lemma 3 and let  $q = \pi\sigma^{-2/3}$  in the DANLM algorithm. Then,*

$$\begin{aligned} \mathbb{P}(\lambda(P_1) - \lambda(R_1) > \epsilon) &\leq O\left(\frac{\sigma^8}{\epsilon}\right), \\ \mathbb{P}(\lambda(P_4) - \lambda(R_2) > \epsilon) &\leq O\left(\frac{\sigma^{22/3}}{\epsilon}\right). \end{aligned}$$

*Proof.* We use the technique we developed in Lemma 3 for the proof. Consider a point  $(u, v) \in P_1$  and let  $\theta^* \triangleq \arctan(h'(t_1))$  and  $\hat{\theta}$  represent an angle

from  $\mathcal{O}$  that is closest to  $\theta^*$ . Then,

$$\begin{aligned}
& \mathbb{P} \left( d_{\mathcal{A}}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&= \mathbb{P} \left( \min_{\theta' \in \mathcal{O}} d_{\theta', \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&\geq \mathbb{P} \left( d_{\hat{\theta}, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&\stackrel{(d)}{\geq} 1 - e^{-\frac{n_s n_\ell \tau_\sigma^2}{4}} = 1 - O(\sigma^8), \tag{27}
\end{aligned}$$

where Inequality (d) is a direct consequence of Lemma 2. Again as in the proof of Lemma 3 we conclude that  $\mathbb{E}(\lambda(P_1) - \lambda(R_1)) = O(\sigma^8)$ . Employing this expression in Markov's Inequality leads us to

$$\mathbb{P}(\lambda(P_1) - \lambda(R_1) > \epsilon) \leq O\left(\frac{\sigma^8}{\epsilon}\right).$$

We now prove the second part of this lemma. Consider a point  $(u, v) \in P_4$ . Then,

$$\begin{aligned}
& \mathbb{P} \left( d_{\mathcal{A}}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&= \mathbb{P} \left( \min_{\theta' \in \mathcal{O}} d_{\theta', \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&\leq \sum_{i=1}^q \mathbb{P} \left( d_{\theta_i, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&\leq O(\sigma^{22/3}).
\end{aligned}$$

Therefore, we conclude that  $\mathbb{E}(\lambda(P_4) - \lambda(R_2)) = O(\sigma^{22/3})$ . Combining this with Markov's Inequality establishes the second part of 6.  $\square$

Define the event  $\mathcal{F}$  as

$$\mathcal{F} \triangleq \{\lambda(P_1) - \lambda(R_1) < \sigma^2\} \cap \{\lambda(P_4) - \lambda(R_2) < \sigma^2\}.$$

The risk of the OANLM estimator at  $(t_1, t_2) \in P_1$  is then given by

$$\mathbb{E} \left( \frac{\int_S w^{AN} X dudv}{\int_S w dudv} \right)^2.$$

Define  $U = \left( \frac{\int_S w^{AN} X dudv}{\int_S w dudv} \right)^2$ . We have

$$\mathbb{E}(U) = \mathbb{E}(U \mid \mathcal{F})\mathbb{P}(\mathcal{F}) + \mathbb{E}(U \mid \mathcal{F}^c)\mathbb{P}(\mathcal{F}^c) \leq \mathbb{E}(U \mid \mathcal{F})\mathbb{P}(\mathcal{F}) + \mathbb{P}(\mathcal{F}^c). \quad (28)$$

Lemma 6 confirms that  $\mathbb{P}(\mathcal{F}^c) = O(\sigma^{16/3})$ . Therefore, we should obtain an upper bound for the term  $\mathbb{E}(U \mid \mathcal{F})\mathbb{P}(\mathcal{F})$ . This step of the proof is the same as the proof of Case I in Theorem 2, and therefore we do not repeat it here.

The more challenging step of the proof is to obtain an upper bound for the risk of DANLM for pixels  $(t_1, t_2) \in P_2$ .

Case II:  $(t_1, t_2) \in P_2$ . In this case we again start with defining the following two sets:

$$\begin{aligned} L_1 &\triangleq \{(u, v) \in P_1 \mid w_{u,v}^D = 1\}, \\ L_2 &\triangleq \{(u, v) \in P_4 \mid w_{u,v}^D = 0\}. \end{aligned}$$

As before, our objective is to show that most of the pixels in  $P_1$  contribute in the final estimate of  $DANLM$  and most of the pixels in  $P_4$  do not contribute. The following lemma formalizes this intuition.

**Lemma 7.** *Let  $\delta_s, \delta_\ell, n_s, n_\ell, \tau_\sigma$  be as defined in Lemma 3, and let  $q = \pi\sigma^{-2/3}$ . Then,*

$$\mathbb{P}(\lambda(P_1) - \lambda(L_1) > \epsilon) \leq \frac{\pi\sigma^8}{\epsilon}.$$

*Proof.* We first prove that for some  $\theta \in \mathcal{O}$ ,  $I_{\theta, \delta_s, \delta_\ell}(t_1, t_2)$  does not intersect the edge. Suppose that  $\theta^*$  is such that  $\tan(\theta^*) = h'(t_1)$ , and consider  $\hat{\theta} = \arg \min_{\theta \in \mathcal{O}} |\theta - \theta^*|$ . To ensure that the neighborhood  $I_{\hat{\theta}, \delta_s, \delta_\ell}(t_1, t_2)$  does not intersect with the edge, we need to have

$$\begin{aligned} t_2 + \tan(\hat{\theta})(t' - t_1) - \delta_s &> h(t_1) + h'(t_1)(t' - t_1) \\ &+ \frac{1}{2}h''(t'') \left( t' - \frac{i}{n} \right)^2 \end{aligned} \quad (29)$$



for all  $t' \in [t_1, t_1 + \delta_\ell)$ . Also,  $t'' \in [t_1, t']$ . According to the mean value theorem we have  $\tan(\hat{\theta}) = \tan(\theta^*) + \frac{1}{1+\tan^2(\theta')} \Delta\theta$ , where  $\Delta\theta = \hat{\theta} - \theta^*$ , and  $\theta' \in [-\pi/4, \pi/4]$ , since in our Horizon model we have assumed  $|h'(t_1)| \leq 1$ . Furthermore,  $|h''(t'')| \leq C_2$  and  $\Delta\theta \leq \sigma^{2/3} \leq \frac{\pi}{2}\delta_\ell$ . Therefore, (29) will be satisfied if

$$t_2 > h(t_1) + \frac{\pi\delta_\ell}{2}\delta_\ell + \delta_s + C_2\delta_\ell^2.$$

This constraint holds for all pixels in the region  $P_2$ . Therefore, the neighborhood  $I_{\hat{\theta}, \delta_s, \delta_\ell}(t_1, t_2)$  lies completely in the region above the  $h$ . Therefore,

$$\begin{aligned} & \mathbb{P} \left( d_{\mathcal{A}}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\ &= \mathbb{P} \left( \min_{\theta' \in \mathcal{O}} d_{\theta', \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\ &\geq \mathbb{P} \left( d_{\hat{\theta}, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\ &\stackrel{(e)}{\geq} 1 - e^{-\frac{n_s n_\ell t^2}{4}} = 1 - O(\sigma^8). \end{aligned} \tag{30}$$

where the Inequality (e) is a consequence of 2. By repeating the same argument as the argument in the proof of Lemma 3 we conclude that  $\mathbb{E}(\lambda(P_1) - \lambda(L_1)) = O(\sigma^8)$ . Employing this expression in Markov's Inequality leads us to

$$\mathbb{P}(\lambda(P_1) - \lambda(L_1) > \epsilon) \leq O\left(\frac{\sigma^8}{\epsilon}\right).$$

□

**Lemma 8.** *Let  $\delta_s, \delta_\ell, n_s, n_\ell, \tau_\sigma$  be as defined in Lemma 3, and let  $q = \pi\sigma^{-2/3}$ . Then,*

$$\mathbb{P}(\lambda(P_4) - \lambda(L_2) > \epsilon) \leq O\left(\frac{\sigma^{10/3}}{\epsilon}\right).$$

*Proof.* We first prove that at least half of the area of  $I_{\theta, \delta_s, \delta_\ell}(u, v)$  is located below the edge. For notational simplicity we assume that  $h'(t_1) = 0$ . Consider Figure 14, in which the neighborhood has an arbitrary orientation. Let  $w, v$  be as defined in this figure. We then have that

$$v = \frac{d}{\cos \theta} - \frac{Cw^2}{\cos \theta} - \delta_s \tan \theta = \frac{d - Cw^2 - \delta_s \sin \theta}{\cos \theta} \geq \frac{d - Cw^2 - \delta_s}{\cos \theta}, \tag{31}$$

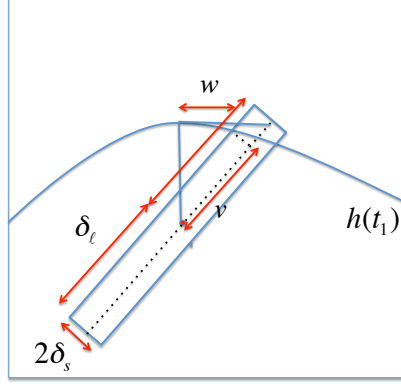


Figure 14: Explanation of the neighborhood rotation in Lemma 8.

where  $d = t_2 - h(t_1)$ . Suppose  $d$  is such that, at the correct angle  $d - Cw^2 - \delta_s > 0$ , i.e.,  $d > (C + 1)\sigma^{4/3}$ . According to (31),  $v > 0$ , and hence more than half of the area of the neighborhood is in the white region (below the edge contour). Therefore, the number of pixels in this region is  $4 \log^2 \sigma \pm o(\log^2 \sigma)$ . Consider a point  $(u, v) \in P_4$ . Then,

$$\begin{aligned}
& \mathbb{P} \left( d_{\mathcal{A}}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&= \mathbb{P} \left( \min_{\theta' \in \mathcal{O}} d_{\theta', \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&\leq \sum_{i=1}^q \mathbb{P} \left( d_{\theta_i, \delta_s, \delta_\ell}^2(dY(t_1, t_2), dY(u, v)) - \frac{2n_s n_\ell \sigma^2}{\tau_s \tau_\ell} \leq \tau_\sigma \right) \\
&\stackrel{(g)}{=} O(\sigma^{10/3}). \tag{32}
\end{aligned}$$

To derive Inequality (g) we have used the fact that  $\Theta(2\delta_\ell \delta_s)$  of the pixels in each neighborhood are equal to 0. To consider the worst case we assumed the rest of the values are equal to 1. Using the same technique as in the proof of Lemma 6 in combination with (32) we obtain

$$\mathbb{P}(\lambda(P_4) - \lambda(L_2) > \epsilon) \leq O\left(\frac{\sigma^{10/3}}{\epsilon}\right).$$

□

Based on Lemmas 7 and 8 we can calculate the risk of DANLM for  $(t_1, t_2) \in P_2$ . Define the event  $\mathcal{G}$  as

$$\mathcal{G} \triangleq \{\lambda(P_1) - \lambda(L_1) < \sigma^2\} \cap \{\lambda(P_4) - \lambda(L_2) < \sigma^2\}.$$

The risk of the OANLM estimator at  $(t_1, t_2) \in P_2$  is then given by

$$\mathbb{E} \left( \frac{\int_S w^{AN} X dudv}{\int_S w dudv} \right)^2.$$

Define  $U = \left( \frac{\int_S w^{AN} X dudv}{\int_S w dudv} \right)^2$ . We have

$$\mathbb{E}(U) = \mathbb{E}(U \mid \mathcal{G})\mathbb{P}(\mathcal{G}) + \mathbb{E}(U \mid \mathcal{G}^c)\mathbb{P}(\mathcal{G}^c) \leq \mathbb{E}(U \mid \mathcal{G})\mathbb{P}(\mathcal{G}) + \mathbb{P}(\mathcal{G}^c). \quad (33)$$

Lemmas 7 and 8 imply that  $\mathbb{P}(\mathcal{G}^c) = O(\sigma^{4/3})$ . Furthermore, following along the same lines of case I, we can prove  $\mathbb{E}(U \mid \mathcal{G})\mathbb{P}(\mathcal{G}) = O(\sigma^{4/3} |\log(\sigma)|^{4/3})$ .

Case III:  $(t_1, t_2) \in S_2$ . The proof of this case is identical to that of Case II in the proof of Theorem 2.

Case IV:  $(t_1, t_2) \in P_3 \cup P_4$ . The proof of this case is similar to that for the regions  $P_1$  and  $P_2$  in Cases I and II and therefore is skipped here.

Finally, combining the upper bounds obtained in Cases I, II, III, and IV completes the proof of the theorem.

## 8. Acknowledgements

This work was supported by the grants NSF CCF-0431150, CCF-0926127, and CCF-1117939; DARPA/ONR N66001-11-C-4092 and N66001-11-1-4090; ONR N00014-08-1-1112, N00014-10-1-0989, and N00014-11-1-0714; AFOSR FA9550-09-1-0432; ARO MURI W911NF-07-1-0185 and W911NF-09-1-0383; and the TI Leadership University Program.

## Appendix A. Minimax risk of the mean filter

In this appendix, we obtain the decay rate of the risk of the mean filter for the Horizon class of images. Our proof is similar to the proof in [4]. Nevertheless, we repeat the proof here for the sake of completeness and since

our continuous framework is slightly different from the discrete framework in [4]. See [41] for the generalization of this result.

The classical mean filter estimates the image via

$$\hat{f}^{\text{MF}}(t_1, t_2) = \frac{1}{\Delta^2} \int_{\tau_1 = -\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau_2 = -\frac{\Delta}{2}}^{\frac{\Delta}{2}} dY(t_1 - \tau_1, t_2 - \tau_2),$$

where  $\Delta$  specifies the size of the window on which averaging takes place.

**Theorem 5** ([4]). *If  $\hat{f}^{\text{MF}}$  is the estimate of the mean filter, then*

$$\inf_{\Delta} \sup_{f \in \mathcal{H}^\alpha(C)} R(f, \hat{f}^{\text{MF}}) = \Theta(\sigma^{2/3}).$$

*Proof.* We first derive a lower bound for  $R(f, \hat{f}^{\text{MF}})$ . Consider a function  $f_h(t_1, t_2)$  with  $h(t_1) = 1/2$  (recall Figure 2) and define

$$Q_\Delta = \{(t_1, t_2) \mid 1/2 - \Delta < t_2 < 1/2 + \Delta\}.$$

It is straightforward to confirm that if  $(t_1, t_2) \in Q_{\Delta/4}$ , then  $|f_h(t_1, t_2) - \mathbb{E}\hat{f}^{\text{MF}}(t_1, t_2)| > \frac{1}{4}$ . Therefore, if  $\text{Bias}(\hat{f}^{\text{MF}})$  denotes the bias of the mean filter estimator, then we have

$$\begin{aligned} \text{Bias}^2(\hat{f}^{\text{MF}}) &= \int_S |f_h(t_1, t_2) - \mathbb{E}\hat{f}^{\text{MF}}(t_1, t_2)|^2 dt_1 dt_2 \\ &\geq \int_{Q_{\Delta/4}} |f_h(t_1, t_2) - \mathbb{E}\hat{f}^{\text{MF}}(t_1, t_2)|^2 dt_1 dt_2 \geq \frac{1}{32} \Delta. \end{aligned} \quad (\text{A.1})$$

Now consider a point  $(t_1, t_2) \in S \setminus Q_\Delta$ . We have

$$\begin{aligned} &\mathbb{E}(\hat{f}^{\text{MF}}(t_1, t_1) - \mathbb{E}\hat{f}^{\text{MF}}(t_1, t_2))^2 \\ &= \int_{\tau_1 = -\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau_2 = -\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau'_1 = -\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau'_2 = -\frac{\Delta}{2}}^{\frac{\Delta}{2}} \frac{\mathbb{E}(dW(t_1 - \tau_1, t_2 - \tau_2) dW(t_1 - \tau'_1, t_2 - \tau'_2))}{\Delta^4} \\ &= \frac{\sigma^2}{\Delta^2} \end{aligned}$$

Therefore, if  $\text{Var}(\hat{f}^{\text{MF}})$  is the variance of the mean filter estimator, then we have

$$\text{Var}(\hat{f}^{\text{MF}}) \geq \int_{S \setminus Q_{4\Delta}} \mathbb{E}(\hat{f}^{\text{MF}}(t_1, t_1) - \mathbb{E}\hat{f}^{\text{MF}}(t_1, t_2))^2 = \frac{\sigma^2}{\Delta^2} (1 - 4\Delta). \quad (\text{A.2})$$

By combining (A.1) and (A.2) we obtain the lower bound of  $\frac{\sigma^2}{\Delta^2} + \frac{1}{32}\Delta - 4\frac{\sigma^2}{\Delta}$  for the risk of the mean filter estimator. If we minimize this lower bound over  $\Delta$  then we obtain the lower bound of  $\Theta(\sigma^{2/3})$  for the risk.

Now, we derive an upper bound for the risk. Define the region

$$R_\Delta = \{(t_1, t_2) \mid h(t_1) - \Delta \leq t_2 \leq h(t_1) + \Delta\}.$$

It is straightforward to confirm that, if  $(t_1, t_2) \in S \setminus R_\Delta$ , then the  $\Delta$ -neighborhood of this point does not intersect the edge contour. Therefore, the bias of the estimator over this region is zero and the variance is  $\frac{\sigma^2}{\Delta^2}$ . If we bound the risk of the points in the  $R_\Delta$  region by 1, we obtain the following upper bound for the risk,

$$R(f, \hat{f}^{\text{MF}}) \leq \frac{\sigma^2}{\Delta^2} + 2\Delta.$$

Again by optimizing the upper bound over  $\Delta$  we obtain the upper bound of  $\Theta(\sigma^{2/3})$ . This completes the proof.  $\square$

## Appendix B. Minimax risk of NLM

Corollary 1 states the following upper bound for the risk of NLM:

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^N) = O(\sigma |\log \sigma|).$$

In this section we prove that the risk of NLM is lower bounded by  $\Omega(\sigma)$ . The proof is similar to the proof of Theorem 5 in [41]; we consider the function  $f_h(t_1, t_2)$  for  $h(t_1) = 1/2$  and prove that the bias of the NLM on  $(t_1, t_2) \in Q_{\frac{\delta}{n}}$  is  $\Theta(1)$  for any choice of the threshold parameter. However, in the continuous setting considered here, the steps are more challenging. Following [41] we consider the semi-oracle NLM algorithm (SNLM). The semi-oracle  $\delta$ -distance is defined as

$$\begin{aligned} & \tilde{d}_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) \\ &= \frac{1}{n^2 - 1} (\|\mathbf{x}_{t_1, t_2}^\delta - \mathbf{y}_{s_1, s_2}^\delta\|_2^2 - |\mathbf{x}_{t_1, t_2}^\delta(0, 0) - \mathbf{y}_{s_1, s_2}^\delta(0, 0)|^2), \end{aligned}$$

where

$$\mathbf{x}_{t_1, t_2}^\delta(j_1, j_2) = \int_{(s_1, s_2) \in I_{\frac{\delta}{n}}^{j_1, j_2}} f(s_1, s_2) ds_1 ds_2.$$

SNLM then estimates the weights according to

$$w_{t_1, t_2}^S(s_1, s_2) = \begin{cases} 1 & \text{if } \tilde{d}_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) \leq \frac{n^2 \sigma^2}{\delta^2} + \tau_\sigma, \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that the distance estimates of the SNLM algorithm are more accurate than those of NLM algorithm. Hence it outperforms NLM, and a lower bound that holds for SNLM will hold for NLM as well.

We make the following mild assumptions on the parameters of SNLM.

A1: The window size  $\delta \rightarrow 0$  as  $\sigma \rightarrow 0$ .

A2:  $\frac{\delta^2}{n^2} = \Omega(\sigma^2)$ . Otherwise  $\mathbb{E} \tilde{d}_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) \rightarrow \infty$  as  $\sigma \rightarrow 0$ .

A3:  $n \rightarrow \infty$  as  $\sigma \rightarrow 0$ . This ensures that, if two points  $(t_1, t_2)$  and  $(s_1, s_2)$  have the same neighborhoods, then  $w_{t_1, t_2}(u_1, u_2) = 1$  with high probability.

A4: If  $\tilde{d}(f(t_1, t_2), f(s_1, s_2)) > 1/4$ , then  $\mathbb{P}(w_{t_1, t_2}^S(s_1, s_2) = 1) = o(\sigma^3)$ .

Again, consider the function  $f_h(t_1, t_2)$  for  $h(t_1) = 1/2$  (recall Figure 2). Let  $(t_1, t_2) \in Q_{\delta/n}$ . For notational simplicity we use  $w(s_1, s_2)$  instead of  $w_{t_1, t_2}^S(s_1, s_2)$ .

**Lemma 9.** *If  $|s_1 - t_1| > \delta/2$  and  $|s'_1 - t_1| > \delta/2$ , then*

$$\mathbb{P}(w_{t_1, t_2}^S(s_1, s_2) = 1) = \mathbb{P}(w_{t_1, t_2}^S(s'_1, s_2) = 1)$$

*for any  $t_1, t_2, s_1, s_2$  and  $s'_1$ .*

The proof is straightforward and hence skipped here. Note that we do not consider the boundary points whose neighborhoods are not subset of  $S = [0, 1]^2$ .

Consider a point  $(t_1, t_2) \in G_{\delta/n}$ .

**Lemma 10.** *If  $|s - t_1| > \delta/2$ , for  $u < \delta/4$ , then we have*

$$\mathbb{P}(w_{t_1, t_2}^S(s, t_2 - u) = 1) = \mathbb{P}(w_{t_1, t_2}^S(s, t_2 + u) = 1)$$

This lemma is a straightforward application of symmetry, and hence we skip the proof.

**Lemma 11.** *Suppose that  $\delta$  and  $\tau_\sigma$  satisfy A1–A4. Then we have*

$$\mathbb{P} \left( \int_{S \setminus Q_{\delta/2}} w(s_1, s_2) ds_1 ds_2 > \sigma^2 \right) = o(\sigma).$$

*Proof.* Define

$$B = \{(s_1, s_2) \in S \setminus Q_{\delta/2} \mid w(s_1, s_2) = 1\}$$

and the event

$$\mathcal{G} = \{\lambda(B) \geq \sigma^2\}.$$

We have

$$\begin{aligned} \mathbb{E}(\lambda(B)) &= \mathbb{E} \int \int_{S \setminus Q_{\delta/2}} I((u, v) \in B) dudv \\ &= \int_{S \setminus Q_{\delta/2}} \mathbb{P}((u, v) \in B) dudv \stackrel{(i)}{=} o(\sigma^3) \end{aligned} \quad (\text{B.1})$$

Equality (i) is due to Assumption (A4); at least quarter of the pixel values in the isotropic neighborhood of any  $(u, v) \in S \setminus Q_{\delta/2}$  are different from the corresponding pixel values of the neighborhood around  $(t_1, t_2) \in Q_{\delta/n}$ . Using the Markov Inequality it is straightforward to show that

$$\begin{aligned} \mathbb{P}(\mathcal{G}) &= \mathbb{P} \left( \int_{S \setminus Q_{\delta/2}} w(u, v) dudv > \sigma^2 \right) \leq \frac{\mathbb{E} \left( \int_{S \setminus Q_{\delta/2}} w(u, v) dudv \right)}{\sigma^2} \\ &= \frac{\mathbb{E} \left( \int_{S \setminus Q_{\delta/2}} I((u, v) \in B) dudv \right)}{\sigma^2} = o(\sigma), \end{aligned} \quad (\text{B.2})$$

where the last equality is the result of (B.1). Finally, if  $\lambda(B) < \sigma^2$ , then

$$\int_{S \setminus Q_{\delta/2}} w(u, v) dudv = O(\sigma^2).$$

This completes the proof.  $\square$

Our next step toward the proof of a lower bound for the risk of NLM algorithm is to show that the bias of the pixels that are in  $Q_{\delta/n}$  is  $\Theta(1)$ . To achieve this goal we show that if  $(t_1, t_2)$  is above the edge and  $(s_1, s_2)$  is below the edge then there is a non-vanishing chance (not converging to zero as  $\sigma \rightarrow 0$ ) that  $(s_1, s_2)$  may contribute to the estimate of  $(t_1, t_2)$ . The following theorem is a formal presentation of this claim.

**Proposition 1.** *Let  $(t_1, t_2), (s_1, s_2) \in Q_{\delta/n}$ . Then there exists  $p_0 > 0$  independent of  $\sigma$  such that for every  $t_n, \delta$  we have*

$$\mathbb{P}(w_{t_1, t_2}^n(u, v) = 1) > p_0.$$

*Proof.* We need to demonstrate that regions above and below the edge contour are included in the NLM algorithm with a constant non-zero probability  $p_0$ . First, we use the definition of the weights of the NLM algorithm to determine the probability that  $w(u, v) = 1$

$$\begin{aligned} \mathbb{P}(w_{t_1, t_2}^n(s_1, s_2) = 1) &= \mathbb{P}\left(\tilde{d}_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) < \frac{n^2\sigma^2}{\delta^2} + \tau_\sigma\right) \\ &= \mathbb{P}\left(\frac{1}{n^2-1} \sum \left(s_{\ell, p}^2 - \frac{n^2\sigma^2}{\delta^2}\right) - \frac{1}{n^2-1} \sum s_{\ell, 0} \leq -\frac{1}{n} + \tau_\sigma\right) \\ &\geq \mathbb{P}\left(\frac{1}{n^2-1} \sum \left(s_{\ell, p}^2 - \frac{n^2\sigma^2}{\delta^2}\right) - \frac{1}{n^2-1} \sum s_{\ell, 0} \leq -\frac{1}{n}\right), \end{aligned}$$

where  $s_{\ell, p} = \mathbf{z}_{s_1, s_2}^{0, \delta, \delta}(\ell, p)$ . Using the Berry-Esseen Central Limit Theorem for independent non-identically distributed random variables [54], we can easily bound this probability away from zero. For more details, see Proposition 1 in [41]. Note that according to Assumption A2 we have  $\frac{n^2\sigma^2}{\delta^2} = O(1)$ .  $\square$

We now consider the weights for the regions above and below the edge contour separately. Therefore, we define

$$\begin{aligned} Q_\Delta^1 &\triangleq \{(t_1, t_2) \mid 1/2 < t_2 < 1/2 + \Delta/4\}, \\ Q_\Delta^2 &\triangleq \{(t_1, t_2) \mid 1/2 - \Delta/4 < t_2 < 1/2\}, \\ \Omega_{\Delta, \delta}^1 &\triangleq \{(t_1, t_2) \mid 1/2 < t_2 < 1/2 + \Delta/4, 0 < t_1 < 2\delta\}, \\ \Omega_{\Delta, \delta}^2 &\triangleq \{(t_1, t_2) \mid 1/2 - \Delta/4 < t_2 < 1/2, 0 < t_1 < 2\delta\}. \end{aligned}$$

Let

$$p_{u, \sigma} \triangleq \mathbb{P}(w_{(t_1, t_2)}^n(u, v) = 1).$$



Note that according to Lemma 9 this probability does not depend on  $v$  and hence we have used the notation  $p_{u,\sigma}$ .

**Lemma 12.** *If  $w(u, v)$  denotes the weights used in the NLM algorithm, then we have*

$$\begin{aligned} \mathbb{P} \left( \left| \int_{Q_\Delta^1} w(u, v) dudv - \int_{Q_\Delta^1} p_{u,\sigma} du \right| > 2\frac{\epsilon}{\delta} + \sqrt{\frac{|\log(\sigma)|}{\delta}} \Delta\delta \right) &= O \left( \frac{\sigma^4}{\epsilon} \right), \\ \mathbb{P} \left( \left| \int_{Q_\Delta^2} w(u, v) dudv - \int_{Q_\Delta^2} p_{u,\sigma} du \right| > 2\frac{\epsilon}{\delta} + \sqrt{\frac{|\log(\sigma)|}{\delta}} \Delta\delta \right) &= O \left( \frac{\sigma^4}{\epsilon} \right). \end{aligned}$$

*Proof.* The weights  $w(u, v)$  over the pixelated neighborhoods can be rewritten in the following way

$$\int_{Q_\Delta^1} w(u, v) dudv = \int_{\Omega_\Delta^1} \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) dudv.$$

Applying the Hoeffding inequality, it is straightforward to confirm that

$$\mathbb{P} \left( \left| \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) - \frac{p_{u,\sigma}}{2\delta} \right| > t \right) \leq 2e^{-2\delta t^2}. \quad (\text{B.3})$$

Now, defining

$$\Gamma_\Delta^1 \triangleq \left\{ (u, v) \in \Omega_\Delta^1 \mid \left| \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) - \frac{p_{u,\sigma}}{2\delta} \right| < t \right\},$$

we see that

$$\begin{aligned} \mathbb{P}(\lambda(\Omega_\Delta^1) - \lambda(\Gamma_\Delta^1) > \epsilon) &= \mathbb{P} \left( \int_{(u,v) \in \Omega_\Delta^1} (1 - \mathbb{I}((u, v) \in \Gamma_\Delta^1)) dudv > \epsilon \right) \\ &\stackrel{(ii)}{\leq} \frac{\left( \mathbb{E} \int_{(u,v) \in \Omega_\Delta^1} (1 - \mathbb{I}((u, v) \in \Gamma_\Delta^1)) dudv \right)}{\epsilon} \\ &= \frac{\left( \int_{(u,v) \in \Omega_\Delta^1} (1 - \mathbb{P}((u, v) \in \Gamma_\Delta^1)) dudv \right)}{\epsilon} \\ &\stackrel{(iii)}{\leq} \frac{2e^{-2\delta t^2}}{\epsilon}, \end{aligned}$$

where (ii) is due to Markov Inequality and Inequality (iii) is due to (B.3). Furthermore,

$$\begin{aligned}
& \left| \int_{\Gamma_{\Delta}^1} \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) dudv - \int_{\Gamma_{\Delta}^1} \frac{p(u, \sigma)}{2\delta} dudv \right| \\
& \leq \int_{\Gamma_{\Delta}^1} \left| \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) - \frac{p(u, \sigma)}{2\delta} \right| dudv \\
& \stackrel{(iv)}{\leq} t\lambda(\Gamma_{\Delta}^1) \leq t\lambda(\Omega_{\Delta}^1) \leq t\Delta\delta,
\end{aligned} \tag{B.4}$$

where Inequality (iv) is due to the definition of  $\Gamma_{\Delta}^1$ . Define the event

$$\mathcal{F} \triangleq \{\lambda(\Omega_{\Delta}^2) - \lambda(\Gamma_{\Delta}^2) < \epsilon\}.$$

Given  $\mathcal{F}$  holds, we conclude  $\lambda(\Omega_{\Delta}^2) - \lambda(\Gamma_{\Delta}^2) < \epsilon$  and hence

$$\begin{aligned}
& \left| \int_{Q_{\Delta}^1} w(u, v) dudv - \int_{Q_{\Delta}^1} p(u, \sigma) dudv \right| \\
& \leq \left| \int_{Q_{\Delta}^1} \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) dudv - \int_{Q_{\Delta}^1} \frac{p(u, \sigma)}{2\delta} dudv \right| \\
& \leq \left| \int_{Q_{\Delta}^1} \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) dudv - \int_{\Gamma_{\Delta}^1} \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) dudv \right| \\
& \quad + \left| \int_{\Gamma_{\Delta}^1} \sum_{k=1}^{\frac{1}{2\delta}} w(u + 2k\delta, v) dudv - \int_{\Gamma_{\Delta}^1} \frac{p(u, \sigma)}{2\delta} dudv \right| \\
& \quad + \left| \int_{\Gamma_{\Delta}^1} \frac{p(u, \sigma)}{\delta} dudv - \int_{Q_{\Delta}^1} \frac{p(u, \sigma)}{\delta} dudv \right| \\
& \stackrel{(v)}{\leq} 2\frac{\epsilon}{\delta} + t\Delta\delta,
\end{aligned}$$

where Inequality (v) is due to (B.4) and the fact that we have assumed  $\mathcal{F}$  holds. Setting  $t = \sqrt{\frac{|\log \sigma|}{\delta}}$  completes the proof.  $\square$

**Theorem 6.** Suppose that  $\delta$ ,  $\tau_\sigma$  and  $n$  satisfy Assumptions A1–A4. Then the risk of SNLM is

$$\inf_{\delta_n, t_n} \sup_{f \in H^\alpha(C)} R(f, \hat{f}^S) = \Omega(\sigma).$$

*Proof.* Let  $(t_1, t_2) \in \Omega_{\delta/n}^2$ . Consider the following two definitions from the proofs of Lemma 11 and Lemma 12:

$$\mathcal{G} \triangleq \{\lambda(B) \geq \sigma^2\},$$

where

$$B = \{(s_1, s_2) \in S \setminus Q_{\delta/2} \mid w(s_1, s_2) = 1\},$$

and

$$\mathcal{F} \triangleq \{\lambda(\Omega_{\delta/2}^2) - \lambda(\Gamma_{\delta/2}^2) < \sigma^2\}.$$

Note that according to Lemmas 11 and 12

$$\mathbb{P}(\mathcal{F}^c \cup \mathcal{G}^c) = o(\sigma).$$

We will calculate a lower bound for the risk of NLM. We have

$$\mathbb{E} \left( f(t_1, t_2) - \frac{\int_S w(u, v) X(u, v) dudv}{\int_S w(u, v) dudv} \right)^2 \geq \left( \mathbb{E} \left( \frac{\int_S w(u, v) X(u, v) dudv}{\int_S w(u, v) dudv} \right) \right)^2.$$

which leads us to towards the lower bound

$$\begin{aligned}
\mathbb{E} \left( \frac{\int_S w(u, v) X(u, v) dudv}{\int_S w(u, v) dudv} \right) &\geq \mathbb{E} \left( \frac{\int_S w(u, v) X(u, v) dudv}{\int_{Q_{\delta/2}} w(u, v) dudv} \right) \\
&\geq \mathbb{E} \left( \frac{\int_S w(u, v) X(u, v) dudv}{\int_{Q_{\delta/2}} w(u, v) dudv} \mid \mathcal{F} \cap \mathcal{G} \right) \mathbb{P}(\mathcal{F} \cap \mathcal{G}) \\
&\stackrel{(vi)}{\geq} \mathbb{E} \left( \frac{\int_S w(u, v) X(u, v) dudv}{\int_{Q_{\delta/2}} p_{u, \sigma} dudv - 2\frac{\sigma^2}{\delta} - |\log(\sigma)|\delta^{3/2}} \mid \mathcal{F} \cap \mathcal{G} \right) \mathbb{P}(\mathcal{F} \cap \mathcal{G}) \\
&\geq \mathbb{E} \left( \frac{\int_S w(u, v) X(u, v) dudv}{\int_{Q_{\delta/2}} p_{u, \sigma} dudv - 2\frac{\sigma^2}{\delta} - |\log(\sigma)|\delta^{3/2}} \right) - \mathbb{P}(\mathcal{F}^c \cup \mathcal{G}^c) \\
&= \left( \frac{\int_S w(u, v) f(u, v) dudv}{\int_{Q_{\delta/2}} p_{u, \sigma} dudv - 2\frac{\sigma^2}{\delta} - |\log(\sigma)|\delta^{3/2}} \right) - \mathbb{P}(\mathcal{F}^c \cup \mathcal{G}^c) \\
&\stackrel{(vii)}{\leq} \left( \frac{\int_{Q_{\delta/2}^1} w(u, v) dudv + \sigma^2}{\int_{Q_{\delta/2}} p_{u, \sigma} dudv - 2\frac{\sigma^2}{\delta} - t\delta^2} \right) - \mathbb{P}(\mathcal{F}^c \cup \mathcal{G}^c) \\
&\stackrel{(viii)}{\leq} \left( \frac{\int_{Q_{\delta/2}^1} p_{u, \sigma} dudv + 2\frac{\sigma^2}{\delta} + |\log(\sigma)|\delta^{3/2} + \sigma^2}{\int_{Q_{\delta/2}} p_{u, \sigma} dudv - 2\frac{\sigma^2}{\delta} - t\delta^2} \right) - \mathbb{P}(\mathcal{F}^c \cup \mathcal{G}^c) \\
&\geq \mathbb{E} \left( \frac{\int_{Q_{\delta/2}^1} p_{u, \sigma} dudv + 2\frac{\sigma^2}{\delta} + |\log(\sigma)|\delta^{3/2} + \sigma^2}{\sigma + 2 \int_{Q_{\delta/2}^1} p_{u, \sigma} dudv - 2\frac{\sigma^2}{\delta} - |\log(\sigma)|\delta^{3/2}} \right) - \mathbb{P}(\mathcal{F}^c \cup \mathcal{G}^c). \tag{B.5}
\end{aligned}$$

Inequality (vi) is an immediate consequence of Lemma 12. Inequality (vii) is due to Lemma 11. Finally Inequality (viii) is the result of Lemma 12. The minimum of the last line is achieved when  $\int_{Q_{\delta/4}^1} p_{u, \sigma} dudv$  is minimized. However, we have proved in Proposition 1 that  $p(u) > p_0$  for  $(u, v) \in Q_{\delta/n}^1$ . Therefore,  $\int_{Q_{\delta/4}^1} p_{u, \sigma} dudv$  is lower bounded by  $\Theta(\frac{\delta}{n})$  which is equal to  $\Omega(\sigma)$  according to Assumption A2. When we substitute this optimum value in the lower bound (B.5), we see that the risk over this region is  $\Theta(1)$ . Therefore the bias of the NLM over the entire image is  $\Omega(\frac{\delta}{n})$  or, equivalently,  $\Omega(\sigma)$  according to Assumption A2.  $\square$

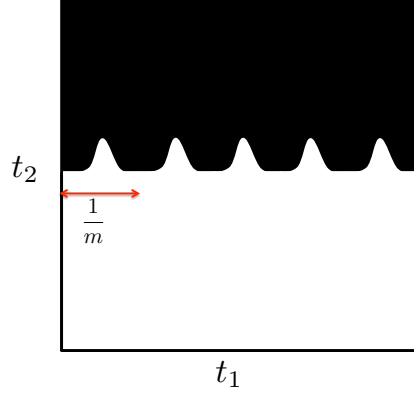


Figure C.15: A sample function from  $\mathcal{F}_m$ . In this figure,  $\zeta_i = 1$  for  $i = 1, 2, \dots, m-1$ .

## Appendix C. Proof of Theorem 1

Here we focus on the case of  $\alpha = 2$  and use the standard technique of hypercube construction to establish the lower bound [10]. Let  $\phi : [0, 1] \rightarrow \mathbb{R}^+ \cup \{0\}$  be a two-times differentiable function with  $\|\frac{d^2\phi}{dt^2}\|_\infty = 1$ ,  $\phi(0) = \phi(1) = 0$ ,  $\frac{d\phi}{dt}\big|_0 = \frac{d\phi}{dt}\big|_1 = 0$ , and  $\frac{d^2\phi}{dt^2}\big|_0 = \frac{d^2\phi}{dt^2}\big|_1 = 0$ . For  $i = 0, 1, \dots, m-1$  define

$$\phi_{i,m} \triangleq \begin{cases} Cm^{-2}\phi(mt - i) & \frac{i}{m} \leq t \leq \frac{i+1}{m}, \\ 0 & \text{otherwise} \end{cases}$$

Consider the function  $f_0 : [0, 1]^2 \rightarrow [0, 1]$ ,  $f_0 \triangleq \mathbf{1}_{\{t_2 < 0.5\}}$ , and define  $\psi_{i,m} \triangleq \mathbf{1}_{\{t_2 \leq \phi_{i,m}(t_1) + 0.5\}} - f_0$ . Finally, define

$$\mathcal{F}_m \triangleq \{f_0 + \sum_{i=1}^m \zeta_i \psi_{i,m}(t_1, t_2), \quad \zeta_i \in \{0, 1\}\}.$$

A sample function of this class is displayed in Figure C.15. Since  $\mathcal{F}_m \subset H^2(C)$ , we have

$$\inf_{\hat{f}} \sup_{f \in H^2(C)} \mathbb{E} \|f - \hat{f}\|_2^2 \geq \inf_{\hat{f}} \sup_{f \in \mathcal{F}_m} \mathbb{E} \|f - \hat{f}\|_2^2. \quad (\text{C.1})$$

The right hand side of (C.1) can be calculated more easily, since we can restrict our attention to the estimators of the form  $f_0 + \sum_{i=1}^m \hat{\zeta}_i \psi_{i,m}$ . This is due to the fact that if  $P_{\mathcal{F}_m}$  is the projection onto  $\mathcal{F}_m$ , then for every  $f \in \mathcal{F}_m$  we have

$$\|P_{\mathcal{F}_m}(\hat{f}) - f\|_2^2 \leq \|P_{\mathcal{F}_m}(\hat{f}) - P_{\mathcal{F}_m}(f)\|_2^2 \leq \|\hat{f} - f\|_2^2.$$

Furthermore for  $f = f_0 + \sum_{i=1}^m \zeta_i \psi_{i,m}(t_1, t_2)$  and  $\hat{f} = f_0 + \sum_{i=1}^m \hat{\zeta}_i \psi_{i,m}(t_1, t_2)$  we have

$$\|f - \hat{f}\|_2^2 = \kappa_m \|\zeta - \hat{\zeta}\|_2^2, \quad (\text{C.2})$$

where  $\kappa_m = \|\psi_{i,m}(t_1, t_2)\|^2$  and satisfies

$$\kappa_m = \|\psi_{i,m}(t_1, t_2)\|^2 = \int_{t_1} \int_{t_2} \psi_{i,m}^2(t_1, t_2) = \int_{t_1} \int_{t_2} \psi_{i,m}(t_1, t_2) = \Theta(m^{-3}).$$

According to (C.2) the original problem reduces to one of estimating  $\zeta$ . Therefore, we reduce the problem to the problem of estimating  $\zeta_1, \dots, \zeta_m$  from the observations  $y_1^{\mathcal{F}}, \dots, y_m^{\mathcal{F}}$  given by

$$y_i^{\mathcal{F}} = \int \psi_{i,m}(t_1, t_2) dY(t_1, t_2) = \|\psi_{i,m}\|_2^2 \zeta_i + \int \sum_i \psi_{i,m} dW(t_1, t_2).$$

We have

$$y_i^{\mathcal{F}} = \kappa_m \zeta_i + w_i^{\mathcal{F}},$$

where  $w_i^{\mathcal{F}} \stackrel{iid}{\sim} N(0, \sigma^2 \kappa_m)$ . Consider the problem of estimating  $\zeta_1$  from the observation  $\zeta_1 + N(0, \sigma^2 / \kappa_m)$  and set  $m$  to the smallest integer for which  $\sigma^2 / \kappa_m \leq 1$ . For this choice of  $m$  the risk of any estimator is lower bounded by a constant  $B$ . Combining this with (C.2) we conclude that minimax risk defined in (C.1) is lower bounded by  $Bm\kappa_m$ . It is straightforward to confirm that  $m = \sigma^{-2/3}$  satisfies the condition  $\sigma^2 / \kappa_m \leq 1$ . Therefore we finish the proof by setting  $m = \sigma^{-2/3}$  and  $\kappa = \sigma^2$ .

## References

- [1] Barbara test image, [http://decsai.ugr.es/~javier/denoise/test\\_images/index.htm](http://decsai.ugr.es/~javier/denoise/test_images/index.htm), 2011.
- [2] Boats test image, USC-SIPI Image Database, 2011.

- [3] Wet-paint test image, <http://www.ece.rice.edu/~wakin/images/>, 2011.
- [4] E. Arias-Castro, D.L. Donoho, Does median filtering truly preserve edges better than linear filtering?, *Ann. Stat.* 37 (2009) 1172–1206.
- [5] E. Arias-Castro, J. Salmon, R. Willett, Oracle inequalities and minimax rates for non-local means and related adaptive kernel-based methods, Submitted to *SIAM J. Imag. Sci.* (2012).
- [6] J. Bigun, G. Granlund, J. Wiklund, Multidimensional orientation estimation with applications to texture analysis and optical flow, *IEEE Trans. Pattern Anal. Machine Intell.* 13 (1991) 775–790.
- [7] T. Brox, O. Kleinschmidt, D. Cremers, Efficient nonlocal means for denoising of textural patterns, *IEEE Trans. Image Processing* 17 (2008) 1083–1092.
- [8] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, *Proc. IEEE Int. Conf. Comp. Vision* (2005).
- [9] E. Candès, D.L. Donoho, Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges, Technical Report, Stanford University, 1999.
- [10] E. Candès, D.L. Donoho, Recovering edges in ill-posed inverse problems: Optimality of curvelet frames, *Ann. Stat.* 3 (2002) 784–842.
- [11] E.J. Candès, D.L. Donoho, Ridgelets: a key to higher-dimensional intermittency?, *Philos. Trans. Roy. Soc. A* 357 (1999) 2495–2509.
- [12] P. Chatterjee, P. Milanfar, Is denoising dead?, *IEEE Trans. Image Processing* 19 (2010) 895–911.
- [13] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, *IEEE Trans. Image Processing* 16 (2007) 2080–2095.
- [14] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, A non-local and shape-adaptive transform-domain collaborative filtering, *Proc. Int. Work. Local Nonlocal Approx. Image Processing* (2008).

- [15] C.A. Deledalle, V. Duval, J. Salmon, Non-local methods with shape-adaptive patches (NLM-SAP), *J. Math. Imag. Vision* 43 (2011) 1–18.
- [16] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Trans. Image Processing* 14 (2005) 2091–2106.
- [17] D.L. Donoho, Wedgelets: Nearly minimax estimation of edges, *Ann. Stat.* 27 (1999) 859 – 897.
- [18] V. Duval, J.F. Aujol, Y. Gousseau, A bias-variance approach for the nonlocal means, *SIAM J. Imag. Sci.* 4 (2011) 760–788.
- [19] G. Easley, D. Labate, W. Lim, Sparse directional image representations using the discrete shearlet transform, *Appl. Comput. Harmon. Anal.* 25 (2008) 25–46.
- [20] K. Egiazarian, V. Katkovnik, L. Astola, Adaptive window size image denoising based on ICI rule, in: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, volume 3, pp. 1869–1872.
- [21] M. Elad, On the origin of the bilateral filter and ways to improve it, *IEEE Trans. Image Processing* 11 (2002) 1141–1151.
- [22] X. Feng, P. Milanfar, Multiscale principal components analysis for image local orientation estimation, in: *Proc. Asilomar Conf. Signals, Systems, and Computers*, volume 1, pp. 478–482.
- [23] G. Gilboa, S. Osher, Nonlocal operators with applications to image processing, *SIAM J. Multiscale Model. Simul.* 7 (2008) 1005–1028.
- [24] S. Grewenig, S. Zimmer, J. Weickert, Rotationally invariant similarity measures for nonlocal image denoising, *J. Vis. Comm. Image Rep.* 22 (2011) 117–130.
- [25] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, *SIAM J. Math. Anal.* 91 (2007) 298–318.
- [26] R. Hummel, R. Moniot, Reconstructions from zero crossings in scale space, *IEEE Trans. Acoust., Speech, Signal Processing* 37 (1989) 2111–2130.



- [27] L. Jacques, L. Duval, C. Chaux, G. Peyrè, A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity, *Signal Processing* 91 (2011) 2699–2730.
- [28] I.M. Johnstone, Gaussian processes: Sequence and multiresolution models, 2011. Unpublished.
- [29] C. Kervrann, J. Boulanger, Optimal spatial adaptation for patch-based image denoising, *IEEE Trans. Image Processing* 15 (2006) 2866–2878.
- [30] R. Kimmel, R. Malladi, N. Sochen, Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images, *Int. J. Computer Vision* 39 (2000) 111–129.
- [31] R. Kimmel, N. Sochen, Orientation diffusion or how to comb a porcupine, *J. Vis. Comm. Image Rep.* 13 (2002) 238–248.
- [32] S. Kindermann, S. Osher, P. Jones, Deblurring and denoising of images by nonlocal functionals, *SIAM J. Multiscale Model. Simul.* 4 (2005) 1091–1115.
- [33] J. Koenderink, The structure of images, *Biological Cybernetics* 50 (1984) 363–370.
- [34] A. Korostelev, A. Tsybakov, *Minimax theory of Image Reconstruction*, Lecture Notes in Statistics, Springer-Verlag, 1993.
- [35] G. Kutyniok, D. Labate, Construction of regular and irregular shearlet frames, *J. Wavelet Theo. Appl.* 1 (2007) 1–10.
- [36] R. Laugesen, N. Weaver, G. Weiss, E. Wilson, A characterization of the higher dimensional groups associated with continuous wavelets, *J. Geom. Anal.* 12 (2002) 89–102.
- [37] E.L. Lehman, G. Casella, *Theory of Point Estimation*, New York: Springer-Verlag, 1998.
- [38] A. Levin, B. Nadler, Natural image denoising: Optimality and inherent bounds, in: *Proc. IEEE Int. Conf. Comp. Vision, and Pattern Recognition*, pp. 2833–2840.

- [39] X. Li, M. Orchard, Edge-directed prediction for lossless compression of natural images, *IEEE Trans. Image Processing* 10 (2001) 813–817.
- [40] M. Mahmoudi, G. Sapiro, Fast image and video denoising via nonlocal means of similar neighborhoods, *IEEE Signal Processing Lett.* 12 (2005) 839–842.
- [41] A. Maleki, M. Narayan, R.G. Baraniuk, Suboptimality of nonlocal means for images with sharp edges, *Appl. Comput. Harmon. Anal.* 33 (2012) 370–387.
- [42] A. Maleki, B. Rajaei, H. Pourreza, Rate-distortion analysis of directional wavelets, *IEEE Trans. Image Processing* 21 (2012) 588–600.
- [43] P. Milanfar, A tour of modern image filtering, To appear in *IEEE Signal Processing Mag.* (2012).
- [44] M. Nagao, T. Matsuyama, Edge preserving smoothing, *Computer Graphics and Image Processing* 9 (1979) 394–407.
- [45] J. Orchard, M. Ebrahimi, A. Wong, Efficient nonlocal-means denoising using the svd, in: *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 1732–1735.
- [46] E.L. Pennec, S. Mallat, Sparse geometric image representations with bandelets, *IEEE Trans. Image Processing* 14 (2005) 423–438.
- [47] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Machine Intell.* 12 (1990) 629–639.
- [48] G. Peyrè, E.L. Pennec, C. Dossal, S. Mallat, Geometrical image estimation with orthogonal bandlet bases, in: *Proc. SPIE Ann. Meeting: Wavelet Apps. Signal Imag. Proc.*, volume 6701, pp. 67010–67022.
- [49] L. Pizarro, P. Mrazek, S. Didas, S. Grewenig, J. Weickert, Generalised nonlocal image smoothing, *Int. J. Computer Vision* 90 (2010) 62–87.
- [50] M. Raphan, E. Simoncelli, An Empirical Bayesian Interpretation and Generalization of Nonlocal Means, Technical Report, Courant Institute of Mathematical Sciences, New York University, 2010.

- [51] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D* 60 (1992) 259–268.
- [52] J. Salmon, Y. Strozzecki, From patches to pixels in non-local methods: Weighted-average reprojection, in: *Proc. IEEE Int. Conference Image Processing (ICIP)*, pp. 1929–1932.
- [53] S.M. Smith, J.M. Brady, SUSAN: A new approach to low level image processing, *Int. J. Computer Vision* 23 (1997) 45–78.
- [54] C. Stein, *Approximate Computation of Expectation*, Institute of Mathematical Statistics, 1986.
- [55] H. Takeda, Kernel regression for image processing and reconstruction, MS.c. thesis, University of California Santa Cruz (2006).
- [56] H. Takeda, S. Farsiu, P. Milanfar, Kernel regression for image processing and reconstruction, *IEEE Trans. Image Processing* 16 (2007) 349–366.
- [57] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: *Proc. Int. Conf. Comp. Vision (ICCV)*, pp. 839–846.
- [58] D. Tschumperlè, Fast anisotropic smoothing of multi-valued images using curvature-preserving pde’s, *Int. J. Computer Vision* 68 (2006) 65–82.
- [59] J.W. Tukey, *Exploratory Data Analysis*, Addison Wesley, 1977.
- [60] D. Van De Ville, M. Kocher, Sure-based non-local means, *IEEE Signal Processing Lett.* 16 (2009) 973–976.
- [61] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, P.L. Dragotti, Directionlets: Anisotropic multi-directional representation with separable filtering, *IEEE Trans. Image Processing* 15 (2006) 1916–1933.
- [62] M. Wakin, J. Romberg, H. Choi, R.G. Baraniuk, Wavelet domain approximation and compression of piecewise smooth images, *IEEE Trans. Image Processing* 15 (2006) 1071–1087.
- [63] R.M. Willett, R.D. Nowak, Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging, *IEEE Trans. Med. Imaging* 22 (2003) 332–350.

- [64] R.M. Willett, R.D. Nowak, Multiscale poisson intensity and density estimation, *IEEE Trans. Info. Theo.* 53 (2007) 3171 –3187.
- [65] A. Wong, P. Fieguth, D. Clausi, A perceptually adaptive approach to image denoising using anisotropic non-local means, in: *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 537–540.
- [66] L. Yaroslavsky, *Digital Image Processing – An Introduction*, Springer Verlag, 1985.